

STUDIES OF PROTEIN FOLDING ON MEMBRANES AND IN CROWDED  
ENVIRONMENTS AND BRIDGING THE RESEARCH-TEACHING GAP IN K-12  
SCIENCE

BY

SHARLENE DENOS

DISSERTATION

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Biophysics and Computational Biology  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Doctoral Committee:

Professor Martin Gruebele, Chair & Director of Research  
Professor Robert Clegg  
Professor Robert Gennis  
Professor Antony Crofts

## **Abstract**

This work deals with three important problems in membrane protein folding studies, namely the preparation and storage of homogeneous small unilamellar vesicles (SUV), the development of an algorithm for selecting soluble trans-membrane helices from known membrane proteins, and the characterization of membrane binding of single surface and trans-membrane helices. I then describe the effects of excluded volume on the stability and kinetics of a stable Lambda Repressor mutant Y22WQ33YA3749G. Finally, I discuss two education projects that aim to bridge the gap between scientific research and K-12 teaching.



*To all the children who have touched my life...*

## **Acknowledgements**

Here at the University of Illinois, we are fortunate to have a strong spirit of collaboration among scientists in addition to a wealth of equipment resources and my research has significantly benefited from this. I would specifically like to acknowledge the laboratory of Dr. Charles Zukowski for the use of their lasers for light scattering measurements and their rheometer for viscosity determination, the Center for Microanalysis of Materials and the Center for Microscopic Imaging for the use of their electron microscopes and the laboratory of Dr. Robert (Bob) Gennis for providing us with access to centrifuges for protein purification and a sonicator for vesicle preparation. The Theoretical and Computational Biophysics Group has generously provided computer time as well as expertise and advice for molecular dynamics simulations. In addition, this work would not be possible without funding from the Center for Physics of Living Cells, the National Science Foundation GK-12 Fellowship grant # DGE 0338215 and MCB 0316925 & 0316925, the NIH Molecular Biophysics Training Grant

I would like to thank my committee members, past and present. Bob Gennis offered many helpful insights at the preliminary exam and also worked with me on a tutorial paper focusing on time resolved infrared techniques. Together with Tony Crofts, Bob also taught a Physical Biochemistry course that was very helpful for developing a foundation for my studies and for passing the Biophysics Qualifying Exam. In addition, I am grateful to Tony Crofts for agreeing at the last minute to serve on my prelim

committee and again for my defense committee. Bob Clegg taught a wonderful spectroscopy course and lab which was both rigorous and accessible to someone with little physics background. He has also been a great mentor to me, helping me to pursue my interests in education outreach. I would also like to thank Emad Tajkhorshid, who served on my preliminary exam and was very helpful.

I would also like to thank Bruce Alberts. Though we have never met, his writings have been very inspiring to me. I am grateful for the work he has done to promote scientists' participation in K-12 education.

I am indebted to the many great minds that have passed through the Gruebele group—Houbi Nguyen, Edgar Larios, Charles Dumont, Marja Engels, Sandra Lee, Dan Weidinger, Praveen Chowdary, Praveen Sundaradevan, Seung Joong Kim, Feng Liu, Krish Sarkar, Apratim Dhar, Simon Ebbinghaus, Tripta Mishra, Josh Ballard, Erin Carmichael, Sumit Ashketar, Greg Scott and Eric Gotkowski—who also contributed substantially to the peptide design project. I cannot thank them enough for their advice and camaraderie over the years.

I would like to thank Cindy Dodds for her tireless work in support of IBS projects, for keeping me on track with graduation requirements and for her friendship over the years.

I would like to thank my family and friends, of which there are too many to name. My husband Adam has been my greatest support, as have my best friends, Dr. Zoreisha Cie,

Gabriela Torres, and Veronica Jacome. My nieces and nephews have been a great inspiration to me, especially Zachary, Nizhone, Tyson, Wesley, Aiyana, Cais, Madison and Jackie. I would also like to thank my parents, the Scotts, the Reynolds, the Denoses, the Tsos, the Duggins, the Gaetas, the Vincents, the Romeros, the Sianis, the Busses, the Dixons, the Dalrymples and all of my colorful enormous family!

Finally, I would like to thank my advisor, Martin Gruebele for his guidance and support during my graduate study. He has supported me in my education interests, not just by letting me have time off from the lab, but by actively participating in teacher workshops, pedagogy seminars and lesson development. He even brought his own children to one of the first outreach events I ever did. I have also been consistently impressed with his patience, integrity and the respect with which he treats others. I have never seen him lose his cool, despite many-thousand dollar mistakes. He embraces the role of a mentor and has coached each of his students according to their own diverse interests. For these reasons, he is deeply admired, not just by members of his own lab, but by virtually all the graduate and undergraduate students who know him.

## Table of Contents

### Section I Studies of Membrane Peptide Binding and Insertion

Introduction.....	1
Chapter 1 A Study of Small Unilamellar Vesicle Morphology and Stability.....	4
Chapter 2 A Design Algorithm for Soluble Membrane-Binding Helical Peptides...	20
Chapter 3 Peptide-Membrane Association Dynamics. ....	47

### Section II Effects of Macromolecular Crowding on Protein Folding and Stability

Introduction. ....	75
Chapter 4 Crowding Studies of a Downhill Folder. ....	81

### Section III Bridging the Gap Between Scientific Research and K-12 Teaching

Introduction.....	97
Chapter 5 Research Tools in the Classroom: Integrating Molecular Visualization and Bioinformatics into the High School Biology Curriculum. ....	100
Chapter 6 Building an Effective K-12 Outreach Program.....	144

### Section IV Supplemental Material

Appendix A Laboratory Protocols.....	163
Appendix B Materials for the UIUC Biophysics K-12 Outreach Program.....	188
Appendix C Molecular Visualization and Bioinformatics Lesson Materials.....	206
Appendix D Computer Codes.....	267

# **Section I**

## **Studies of Membrane Peptide Binding and Insertion**

### **Introduction**

Over one third of genes in organisms from bacteria to humans encode membrane proteins. These proteins are responsible for a vast array of vital cellular functions from the generation of ATP to the regulation of cellular water content. As a consequence, misfolding of membrane proteins is often disastrous for cells and is the known cause of several diseases, including cystic fibrosis, Alzheimer's and Parkinson's disease. Membrane proteins are also the targets of a large number of drugs and toxins and are responsible, in part, for their uptake, metabolism, and clearance. Despite their importance, studies of membrane protein folding have lagged far behind those of globular proteins, primarily because their hydrophobicity makes them very difficult to study. In particular, many scientists would like to characterize the dynamics and mechanistic details of the membrane-mediated folding process. An easier way to accomplish this is to study individual trans-membrane helices from a larger membrane protein. However, such studies still require a water-membrane partition coefficient that is of an intermediate value, so that the equilibrium can be perturbed under an accessible set of conditions. Few such peptides exist and so the current generation of peptide binding, insertion and folding experiments has focused on venomous peptides or antimicrobial peptides that normally induce membrane rupture. Examples include computational and experimental studies of mellitin,<sup>1;2</sup> and comprehensive kinetic studies by Gai and coworkers of several peptides.<sup>3;</sup><sup>4;5</sup> In addition, success has been achieved with designed sequences, such as the surface-

binding peptide helix-5,<sup>6</sup> or TMX-3, whose thermodynamics and insertion kinetics have been studied in detail by White and coworkers<sup>7</sup>. Other peptides which display pH-dependent insertion and solubility<sup>2, 8</sup> are amenable to stopped-flow or pH-jump studies, but cannot be studied by temperature jump methods because partitioning cannot be controlled at a single pH. Design and engineering of soluble membrane-binding peptides is difficult because it is difficult to balance the competing effects of hydrophobicity, amphipathicity and charge, which determine equilibrium constants. These equilibrium constants are then exponentially related to free energies, so the feasible region of this space is extremely narrow.

The work presented in this section deals with three important problems in membrane protein folding studies, namely the preparation and storage of homogeneous small unilamellar vesicles (SUV), the development of an algorithm for selecting soluble trans-membrane helices from known membrane proteins, and the characterization of membrane binding of single surface and trans-membrane helices.

## References

1. Constantinescu, I. & Lafleur, M. (2004). Influence of the lipid composition on the kinetics of concerted insertion and folding of melittin in bilayers. *Biochimica et Biophysica Acta* **1667**, 26-37.
2. Tang, J. & Gai, F. (2008). Dissecting the Membrane Binding and Insertion Kinetics of a pHLIP Peptide. *Biochemistry* **47**, 8250-8252.

3. Tang, J., Signarvic, R. S., DeGrado, W. F. & Gai, F. (2007). Role of Helix Nucleation in the Kinetics of Binding of Mastoparan X to Phospholipid Bilayers. *Biochemistry* **46**, 13856-13863.
4. Tucker, M. J., Tang, J. & Gai, F. (2006). Probing the Kinetics of Membrane-Mediated Helix Folding. *Journal of Physical Chemistry B* **110**, 8105-8109.
5. Tang, J., Yin, H., Qiu, J., Tucker, M. J., DeGrado, W. F. & Gai, F. (2009). Using Two Fluorescent Probes to Dissect the Binding, Insertion, and Dimerization Kinetics of a Model Membrane Peptide. *Journal of the American Chemical Society* **131**, 3816-3817.
6. Russel, C. J., King, D. S., Thorgeirsson, T. E. & Shin, Y. K. (1998). De novo design of a peptide which partitions between water and phospholipid bilayers as a monomeric alpha-helix. *Protein Engineering* **11**, 539-547.
7. Ladokhin, A. S. & White, S. H. (2004). Interfacial Folding and Membrane Insertion of a Designed Helical Peptide. *Biochemistry* **43**, 5782-5791.
8. Reshetnyak, Y. K., Segala, M., Andreev, O. A. & Engelman, D. M. (2007). A Monomeric Membrane Peptide that Lives in Three Worlds: In Solution, Attached to, and Inserted across Lipid Bilayers. *Biophysical Journal* **93**, 2363-2372.



## **Chapter 1**

### **A Study of Small Unilamellar Vesicle Morphology and Stability**

#### **1.1 Introduction**

Membrane binding and insertion studies require a well-characterized model membrane. Small unilamellar vesicles (SUV) are ideal for these studies, since they produce less scatter and are amenable to different types of measurements (i.e. near UV circular dichroism and fluorescence measurements at low concentrations of protein). SUVs also offer a low peptide-vesicle ratio, which is important to maintain due to the possibility of membrane-catalyzed peptide association or aggregation. The latter is also interesting to study, but it is important to have control over the extent of such peptide association on the membrane.

Despite their attractive qualities, however, SUVs are not equilibrium structures and due to their high radius of curvature, they are known to fuse, aggregate, and undergo phase transitions over long term storage and at high temperatures<sup>1</sup>. In order to have confidence in membrane-binding and insertion studies, the physical characteristics of the vesicle system used must be well understood.

To achieve this, I study vesicle morphology using transmission electron microscopy; size and homogeneity using dynamic light scattering; and concentration using the total phosphorous assay<sup>2</sup> for samples stored between 1 and 22 days at 4°C. Also, to approximate the thermal stress on samples during folding melts and temperature jumps, I performed dynamic light scattering measurements after 30 minute incubations at 10, 40

and 70°C at each time point. Finally, the presence of phase transitions, such as the  $L_\alpha$  to HII (inverted hexagonal) transition common at high temperatures for many lipids<sup>1</sup> was determined using differential scanning calorimetry. In addition to this stability study, the effect of four variations of the conventional SUV preparation method<sup>3</sup> on the characteristics of freshly prepared vesicles was also compared.

## 1.2 Materials and Methods

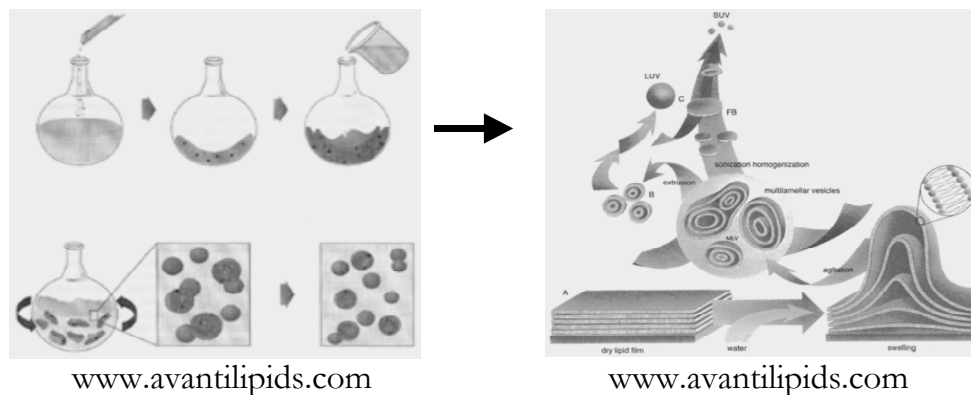
**Vesicle preparation** Small unilamellar vesicles (SUVs) were prepared using lipids purchased from Avanti Polar Lipids (Alabaster, AL). Neutral vesicles are prepared with palmitoyl-oleoyl phosphatidylcholine (POPC), negatively charged vesicles are prepared with 5% palmitoyl-oleoyl phosphatidylglycerol (POPG) in POPC.

The preparation method used for vesicles in the stability study is the following:

- 1) A thin lipid film was obtained on a Pyrex flask by evaporation of lipid in chloroform with dry nitrogen.
- 2) Samples were placed under vacuum at <1 Torr for at least 4 hours prior to reconstitution in buffer.
- 3) The film was resuspended in room temperature, 50 mM sodium phosphate buffer at pH 7 by swirling and mild vortexing.
- 4) Samples were exposed to 5 freeze-thaw cycles alternating a liquid nitrogen bath and a 50°C water bath.
- 5). Samples were sonicated until clear, using a Vibra-Cell Disrupter (Sonics & Materials, Inc., Newtown, CT) fitted with a 5 mm titanium horn and operated at a 60% duty cycle at output control level 5.
- 6) Samples were spun in a TLA 100 Beckman Tabletop Ultracentrifuge (Beckman Coulter, Fullerton, CA), using a TLA 100.3 rotor at 125,000 x g for 2 hours at 10 °C. The top two-thirds of the supernatant was retrieved and the

remaining sample discarded. 7) Samples were extruded using a 50 nm polycarbonate filter in the Avanti Mini Extruder (Avanti Polar Lipids, Alabaster, AL).

**Figure 1.1** Vesicle preparation scheme.



Steps 2, 3, 6 and 7 represent modifications to the conventional SUV preparation method<sup>3</sup>. To determine the effect of these steps on vesicle characteristics, I prepared 4 additional samples in which 1) Chloroform removal under vacuum (step 2 above) is omitted, 2) Alternating freeze-thaw cycles (step 4 above) is omitted, 3) Both ultracentrifugation and extrusion (steps 6 and 7 above) are omitted, and 4) Extrusion (step 7) only is omitted.

### **Total Phosphorous Assay (TPA).**

TPA was used to determine the final phospholipid concentration after filtration and centrifugation steps, which remove a substantial portion of lipid. I used a modified Fiske-Subarrow method<sup>2</sup>, where 4:1 sulfuric to perchloric acid is used to digest the phospholipid to inorganic phosphorous and 2.5% ammonium molybdate in combination with 10% ascorbic acid is used to develop the color. 13 x 188 mm borosilicate tubes

were covered with tinfoil during the hydrolysis step to avoid cross-contamination of these shorter tubes. Reactions are monitored by their absorbance at 820nm against a water blank. Standards are assayed in duplicate, samples in triplicate. The final vesicle concentration is calculated by dividing the number of lipids per vesicle (of the diameter obtained in DLS measurements) by the phospholipid concentration determined from the TPA. This is done assuming an area of  $64\text{\AA}^2$  per lipid and a 3.9 nm bilayer thickness which corresponds to  $\approx 21,000$  lipids per 50 nm vesicle <sup>4</sup>.

### **Negative stain transmission electron microscopy (TEM).**

TEM was used to qualitatively assess vesicle size, shape and lamellarity. Samples were dried on 300 mesh, formvar and carbon coated copper grids obtained from Electron Microscopy Sciences (Hatfield, PA). Grids were not used more than one month after glow discharge since they become considerably less hydrophilic after this time and are, therefore, inefficient at binding to the sample. To bind sample, grids are floated on top of a 2uL droplet of an approximately 5-10uM vesicle suspension for 15-30 minutes. The sample is then wicked off (not blotted) using Whatman filter paper before floating the grid atop a 2uL drop of 2.5% ammonium molybdate ( $\text{NH}_4\text{Mo}$ ), pH 6.0-6.5 for not longer than 2 minutes. After wicking off the stain solution, samples are allowed to dry at least 30 minutes before they are placed in the microscope. The microscope used is a Phillips CM12 TEM fitted with a Gatan UltraScan 1000 CCD camera. The microscope is operated at 120kV and the resolution is 0.34nm, near the diffraction limit at this voltage. The diffraction limited resolution,  $R_d$ , is given by:

$$R_d = \frac{1.23nm}{2 \bullet NA \bullet V} \quad (1)$$

where NA is the numerical aperture of the electromagnetic lens used to focus the electron beam (on the order of  $10^{-5}$ ) and V is the accelerating voltage passed between the Tungsten filament (cathode) and the anode. Digital images are acquired using a Gatan 794 1kx1k Slow Scan CCD and the Gatan Digital Micrograph 3.0 software package. The noise is very low in this setup so that no processing of TEM images is required.

**Dynamic light scattering (DLS)** was performed using the Brookhaven Instruments goniometer system (BI-200SM, Brookhaven Instruments Corp., Holtsville, NY) equipped with a Lexel argon-ion laser (model 95, Cambridge Lasers Lab., Fremont, CA) operating at 514 nm. Temperature control is achieved using a NesLab (Thermo Scientific, Waltham, MA) re-circulating water bath, after incubation for 30 minutes at the given temperature. The coherence of scattered light from the vesicle suspension was measured using a photomultiplier tube placed after a 100nm aperture at 90° from the path of the laser beam and using a Brookhaven Instruments BI-9000AT Digital Autocorrelator. The intensity or second order autocorrelation function is expressed by the Siegert relationship <sup>5</sup>:

$$G^{(2)}(\tau) = B \bullet (1 + f^2 |g^{(1)}(\tau)|^2) \quad (2)$$

where  $G^{(2)}(\tau)$  is the raw measured intensity,  $\tau$  is the delay time, B is the baseline, f is a fit parameter which is a function of the spatial and temporal coherence of laser and detector

optics, and  $g^{(1)}(\tau)$  is the first order, or electric field autocorrelation function. The function  $g^{(1)}(\tau)$  is normalized, whereas  $G^{(2)}(\tau)$  is not. Assuming monodisperse, rigid, globular scatterers,

$$|g^{(1)}(\tau)| = \exp(-\Gamma \bullet \tau) \quad (3)$$

where  $\Gamma$  is the decay rate and  $\tau_r = 1/\Gamma$  is the decay time. The diffusion coefficient,  $D_T$ , is directly related to the decay rate according to:

$$\Gamma = D_T \bullet q^2 \quad (4)$$

where  $q$  is the magnitude of the scattering wave vector, defined as:

$$q = \frac{4\pi \bullet n_0}{\lambda_0} \bullet \sin(\theta / 2) \quad (5)$$

Here,  $n_0$  refers to the refractive index of the solvent,  $\lambda_0$  is the wavelength of the laser in a vacuum, and the scattering angle is given by  $\theta$ . Since the refractive index is a strong function of temperature, the sample compartment is maintained at a constant temperature and is closely monitored with a thermocouple. The mean effective hydrodynamic diameters,  $d_h$ , are then computed from the measured diffusion coefficients using the Stokes-Einstein relation for spherical particles:

$$D_T = \frac{k_B T}{3\pi\eta d_h} \quad (6)$$

where  $\eta$  is the bulk viscosity of the buffer solution,  $k_B$  is Boltzmann's constant, and  $T$  is the absolute temperature. This analysis is strictly valid only for infinitely dilute systems, where the decay can be attributed entirely to translational diffusion of the scatterer's center of mass. At higher concentrations, the square of the wave vector,  $q$ , will be proportional to mutual diffusion, rather than simple self diffusion as described in (4). In practice, however, a lower limit on the concentration of scatterers present is required in order to obtain good photon statistics. The laser is focused down to 100 $\mu$ m at the sample and detection apertures used are typically 200 $\mu$ m. This gives a 1.6 nanoliter detection volume, which will contain nearly 1,000 vesicles even at a 1 picomolar sample concentration. In practice, picomolar suspensions of vesicles are the most dilute samples that will still produce reasonable counts (>300 kiloCounts/second).

A cumulant fit to the autocorrelation function<sup>6</sup> is used to obtain two key parameters, the mean effective diameter,  $d_H$ , and the polydispersity,  $poly$ <sup>7</sup>.  $Poly$  is a measure of the width of the gaussian distribution of scatterer sizes and monodisperse samples are defined as those with  $poly < 0.3$ . The accuracy of results is independent of scatterer size; however this analysis is valid only as long as the size distribution is unimodal and the sample is homogeneous.

**Differential Scanning Calorimetry (DSC)** measurements were performed on a 45mM lipid sample of 45nm anionic SUV. Measurements were made by scanning at 30°C per hour from 10 to 85°C on a Microcal MCS Differential Scanning Calorimeter (Microcal, Northampton, MA). Buffer baseline scans are subtracted from the lipid measurement.

## 1.3 Results

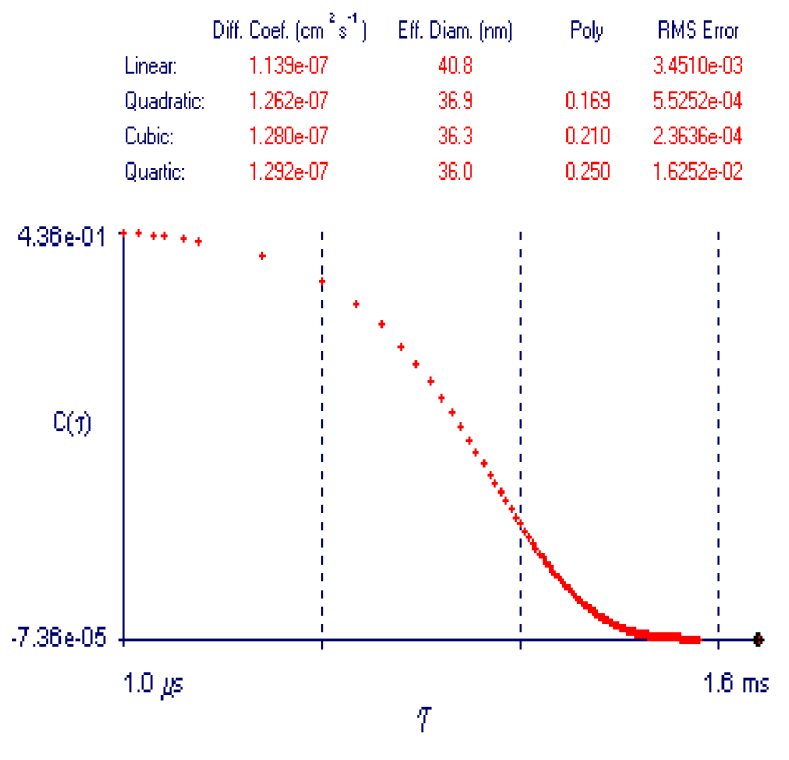
### Stability of Stored Samples.

Figure 1.2 shows the DLS autocorrelation function and cumulant fits for the sample measured at 10°C and 24 hours. The quadratic fit has an RMS error below  $10^{-3}$ ; however, the cubic fit was more often required to achieve similarly low errors for the other samples. Table 1.1 shows the polydispersity and mean effective diameters,  $d_h$ , obtained using the cubic fit for each time and temperature measured in this study. No significant increase is observed in  $d_h$  or size heterogeneity for up to 22 days at room temperature and up to 9 days at temperatures up to 70°C.

TEM images of stored samples are shown in figure 1.3. These images roughly confirm the size measurements from DLS, though this method is much more qualitative and it is not as easy to detect changes in size and heterogeneity with this method. All images were obtained at the same magnification, though the ruler is absent from images 1.3c and d.



**Figure 1.2** DLS results for the 10°C sample after 24 hours of storage at 4°C. Cumulants analysis yields good fits to the autocorrelation function with quadratic or cubic terms.

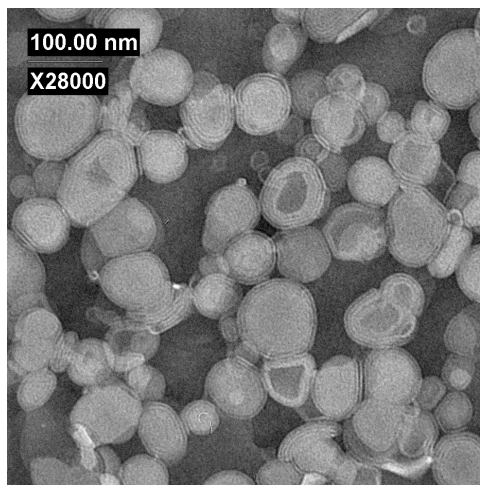


**Table 1.1** DLS results for stored samples at 10°C, 40°C and 70°C. The 22 day time point was measured at 23°C only.

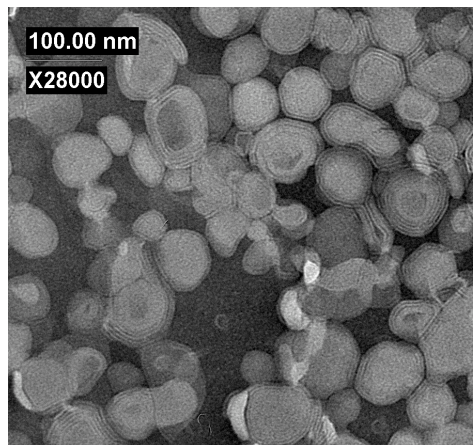
Temperature (C)	Time Since Preparation	<24hrs	~72hrs	9 days	22 days (23°C only)
10	d <sub>h</sub> (nm)	40.1	35.5	37.8	39.9
	Poly	0.210	0.260	0.260	0.240
40	d <sub>h</sub> (nm)	38.5	38.5	38.9	N/A
	Poly	0.228	0.229	0.220	N/A
70	d <sub>h</sub> (nm)	39.0	38.4	38.6	N/A
	Poly	0.227	0.237	0.230	N/A

**Figure 1.3** TEM micrographs of room temperature SUV at **A.** 24hr, **B.** 72hr, **C.** 9 day and **D.** 22 day time points. Micrographs in C and D have the same scale/magnification as those in A and B.

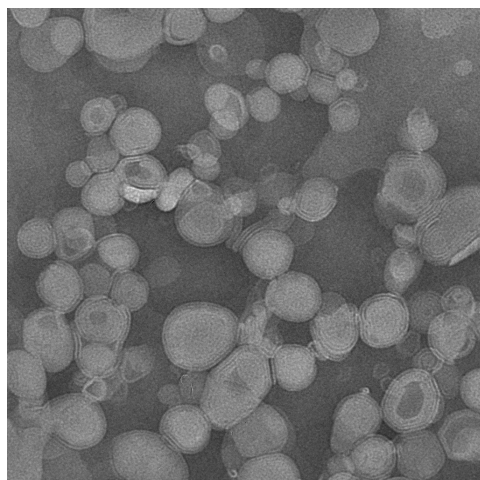
**A.**



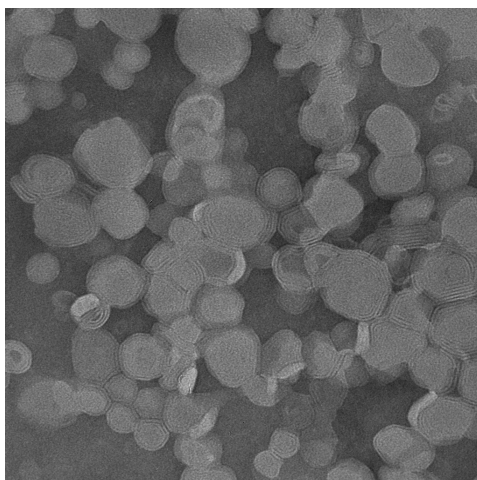
**B.**



**C.**



**D.**



### **Characteristics of Modified Sample Preparations.**

Figure 1.4 and 1.5 show DLS cumulants analysis and TEM micrographs, respectively, of modified SUV preparations. Comparison of these samples with those from the stability study shows that leaving out any protocol steps significantly increases both the heterogeneity and mean diameter of the samples. The vacuum drying step (step 2) has the least effect on these parameters, while ultracentrifugation (step 6) has the greatest effect. TEM micrographs confirm these results, with panel 1.5c showing one of several vesicles found in this sample that were an order of magnitude larger in diameter than stability study samples from figure 1.3. Figure 1.5d, the micrograph from the sample prepared without freeze-thaw cycles, also shows larger vesicles, approximately twice the size of stability study samples in figure 1.3.

### **DSC Results**

DSC results are shown in figure 1.6a and b as excess heat (cal/mol°C) and enthalpy (kCal/mol), respectively. The excess heat increases by approximately 75cal/mol over 75 degrees Celsius. This change is about 1/40<sup>th</sup> of that obtained for the gel to liquid crystalline transition for POPC lipids and about 1/10<sup>th</sup> of the change observed for the less cooperative, liquid crystalline to HII transitions in phospholipids.

**Figure 1.4** DLS results for SUV prepared without selected steps in the protocol. Omitted steps are **A.** Step 2, vacuum drying of lipid film **B.** Step 4, alternating freeze-thaw cycles, **C.** Steps 6 & 7, ultracentrifugation and extrusion **D.** Step 7, extrusion, only. RMS error for the cubic fit in B. was below  $10^{-3}$  as for most samples measured.

**A.**

	Gamma ( $s^{-1}$ )	Eff. Diam. (nm)	Poly	RMS Error
Linear:	4.218e+03	59.5		9.8917e-03
Quadratic:	4.559e+03	55.1	0.119	7.6617e-03
Cubic:	5.133e+03	48.9	0.268	1.6091e+05
Quartic:	5.476e+03	45.8	0.370	3.6421e+05

**B.**

	Gamma ( $s^{-1}$ )	Eff. Diam. (nm)	Poly	RMS Error
Linear:	4.242e+03	58.4		3.2486e-03
Quadratic:	4.952e+03	50.0	0.214	2.2101e-03
Cubic:	5.110e+03	48.5	0.283	1.6649e-04
Quartic:	5.149e+03	48.1	0.308	3.4487e-02

**C.**

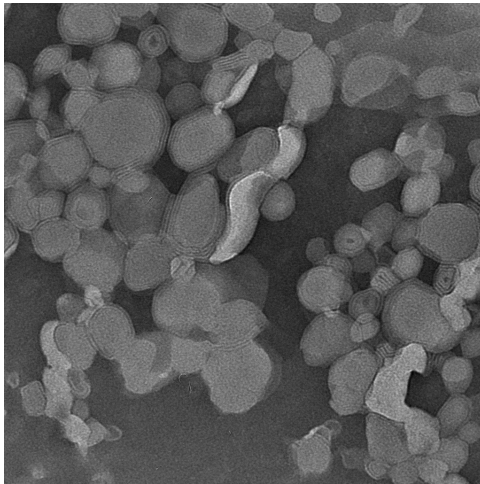
	Gamma ( $s^{-1}$ )	Eff. Diam. (nm)	Poly	RMS Error	A
Linear:	1.965e+03	127.8		2.7836e-02	
Quadratic:	2.660e+03	94.4	0.312	2.9151e+05	
Cubic:	3.044e+03	82.5	0.466	2.0812e-02	
Quartic:	3.475e+03	72.2	0.629	3.1360e+05	

**D.**

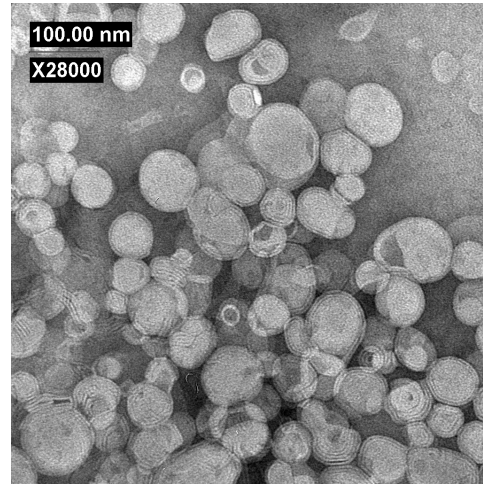
	Gamma ( $s^{-1}$ )	Diff. Coef. ( $cm^2 s^{-1}$ )	Eff. Diam. (nm)	Poly
Linear:	3.671e+03	6.891e-08	67.5	
Quadratic:	4.401e+03	8.262e-08	56.3	0.175
Cubic:	4.951e+03	9.296e-08	50.0	0.324
Quartic:	5.252e+03	9.860e-08	47.2	0.419

**Figure 1.5** TEM results at ambient temperature for SUV prepared without selected steps in the protocol. Omitted steps are **A.** Step 2, vacuum drying of lipid film **B.** Step 4, alternating freeze-thaw cycles, **C.** Steps 6 & 7, ultracentrifugation and extrusion **D.** Step 7, extrusion, only. The micrograph in **A** has the same scale/magnification as the other 3.

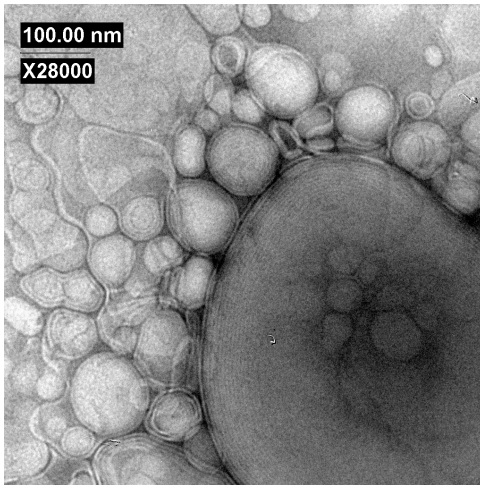
**A.**



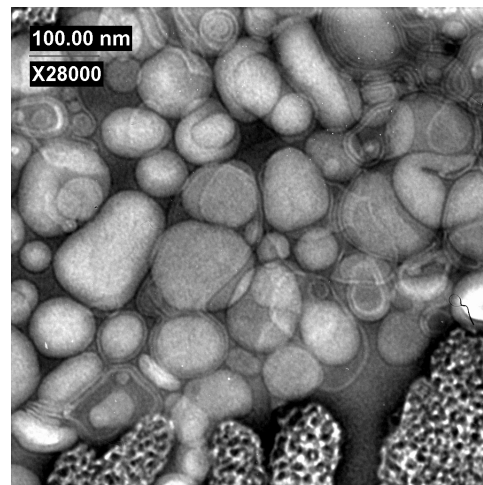
**B.**



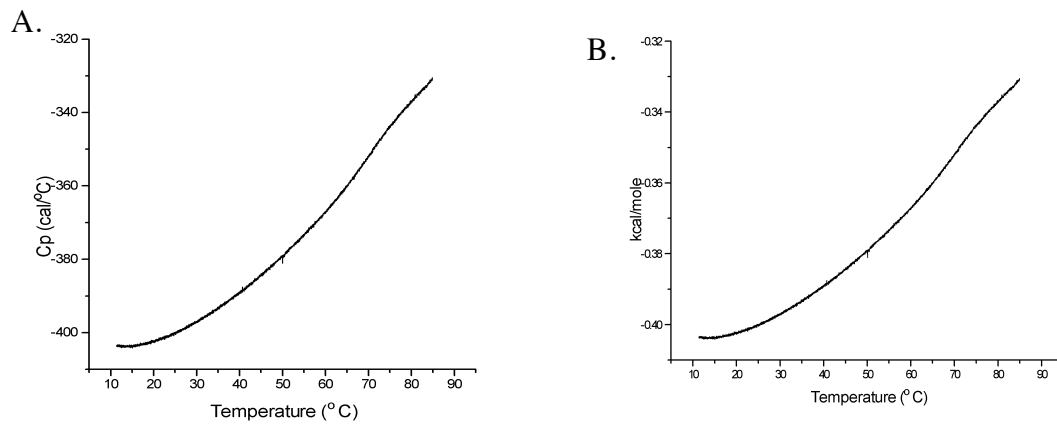
**C.**



**D.**



**Figure 1.6** DSC results for 50nm anionic SUV scanned from 10°C to 85°C. **A.** Results expressed in calories/degree. **B.** Results expressed in kCal/mol.



## 1.4 Discussion

This study establishes a stability window of at least 3 weeks at 4°C storage for SUV prepared according to the full protocol described here. Further, samples stored up to 9 days are stable when incubated for 30 minutes at temperatures as high as 70°C and as low as 10°C. It is likely that this stability window can be significantly extended, since no increase was observed in mean diameter or polydispersity for any sample measured in this stability study. Furthermore, TEM micrographs reveal that these monodisperse samples are mostly unilamellar and morphology does not vary with storage.

Modifications to the conventional SUV preparation presented here are all useful for reducing mean diameter and size heterogeneity. The most important modification, however, is the ultracentrifugation step. Omitting the extrusion step can also result in significantly larger vesicles.

DLS results reveal about a 1cal/mol° excess heat over the temperature range 10 – 85°C. These results show that no phase transition exists for this lipid system in this range.

Overall, I have shown that minor modifications to the conventional SUV preparation method lead to monodisperse vesicles less than 50nm in diameter. Furthermore, these SUV can be stored and used over a wide range of temperatures range without fusion, aggregation or phase transitions.

## 1.5 References

1. Cevc, G. (1993). *The Phospholipid Handbook*, Marcel Dekker Inc., New York.
2. Bartlett, G. R. (1959). Phosphorus Assay in Column Chromatography. *Journal of Biological Chemistry* **234**, 466-468.
3. Barenholz, Y., Gibbes, D., Litman, B. J., Goll, J., Thompson, T. E. & Carlson, F. D. (1977). A Simple Method for the Preparation of Homogeneous Phospholipid Vesicles. *Biochemistry* **16**, 2806-2810.
4. Gullingsrud, J. & Schulten, K. (2004). Lipid Bilayer Pressure Profiles and Mechanosensitive Channel Gating. *Biophysical Journal* **86**, 3496-3509.
5. Johnson, C. S. & Gabriel, D. A. (1994). *Laser Light Scattering*, Dover Publications, Inc. .

6. Koppel, D. E. (1972). Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants. *The Journal of Chemical Physics* **57**, 4814 -4820.
7. Finsy, R. (1994). Particle Sizing by Quasi-Elastic Light Scattering. *Advances in Colloid and Interfacial Science* **52**, 79-143.



## Chapter 2

### A Design Algorithm for Soluble Membrane-Binding Helical Peptides

#### 2.1 Introduction

*In vitro* folding and binding studies of membrane peptides require good solubility of a peptide in both the membrane and aqueous medium, as well as a partition coefficient near unity, so that large signals and large signal changes can be observed when the system is perturbed. The resulting delicate balance between the ability of a peptide to insert into the membrane and its ability to avoid aggregation in the surrounding buffer requires fine-tuning of the peptide's hydrophobicity, amphipathicity and charge. There are many trans-membrane protein helical segments with known high resolution structures to choose from, but it is unclear which subset is likely to bind to membranes on their own, yet remain sufficiently soluble for folding-binding experiments such as thermal titrations or temperature jump kinetics.

In an effort to facilitate the mining of existing membrane protein databases<sup>9</sup> for soluble membrane-binding peptides that undergo easily detectable spectral shifts upon interaction with the membrane, I developed a combined computational and experimental approach involving four major steps. First, I characterized a set of helical membrane peptides experimentally and by sequence-based parameters related to hydrophobicity, amphipathicity and charge. This set was chosen to include both soluble and insoluble peptides. Next, I employed a linear learning algorithm to search for a linear combination of sequence parameters capable of separating this training set optimally into soluble and

insoluble peptides. I obtained an easily computable predictive criterion  $D$ . In the third step, the predictive power of  $D$  was tested by synthesis and characterization of four new trans-membrane helix sequences predicted to be soluble at the micromolar level. All four new peptides were soluble, and three out of four bound to the membrane, as measured by fluorescence wavelength or intensity shifts in the presence of small unilamellar vesicles. In the final step, I predict four additional new soluble membrane-binding peptides.

## **2.2 Materials and Methods**

### **Peptide nomenclature and synthesis.**

Peptides originating from multi-spanning membrane proteins are given the abbreviated protein name followed by a number that indicates which trans-membrane helix (counting from the N-terminus) the peptide corresponds to. *De novo* peptide names are not changed from the original source. All syntheses were performed by Genscript Corp. (Piscataway, NJ), except in the case of TetA-4 and G3PT-11, which were synthesized by the University of Illinois Biotechnology Center (Urbana, IL). All peptides discussed in this study are listed in tables 2.1 through 2.3.

### **Solubility and binding measurements.**

Solubility and binding measurements were performed in 50 mM sodium phosphate buffer. Initial UV/Vis measurements in pH 6, 7, & 8 buffer for each peptide revealed which pH was optimal for further studies. All peptides were studied at pH 7 except for CCOIII-4, which was dissolved in pH 6 buffer. The solubility limit was taken

as the protein concentration left over in the supernatant after centrifugation at 16,000 x g for 20 minutes in a Eppendorf 5415D microcentrifuge (Eppendorf AG, Hamburg, Germany). If no decrease in concentration was observed after this initial spin, an additional centrifugation at 125,000 x g for 2 hours at 10 °C in a Beckman TL-100 ultracentrifuge (Beckman Coulter Inc., Fullerton, California) was performed to obtain the final solubility. Protein concentrations in the supernatant were determined from 280 nm absorbance. Extinction coefficients were calculated by summing all the tryptophan, tyrosine and cysteine residues and assuming a spectral contribution at 280 nm of 5690 M<sup>-1</sup>cm<sup>-1</sup>, 1280 M<sup>-1</sup>cm<sup>-1</sup> and 120 M<sup>-1</sup>cm<sup>-1</sup>, respectively, for each<sup>10</sup>. A peptide is defined as insoluble in the micromolar range if less than 1 µM remains in the supernatant after repeated centrifugation. SUV are prepared as described in chapter 1.

### **Dispersed fluorescence measurements.**

Fluorescence spectra were taken to test whether peptides were capable of SUV binding. All spectra were obtained using a Varian (Palo Alto, CA) Cary Eclipse spectrofluorimeter with a 4-position Peltier thermostatted multicell changer and a PCB-150 circulating water bath. Slit widths of 5 nm were used for both emission and excitation monochromators and dry nitrogen was used to purge the sample chamber while the Peltier heating was controlled using a temperature probe placed directly in the buffer blank cuvette. Background subtraction is performed with buffer alone for the solution melt and SUV in buffer peptide-SUV melts. Spectra were taken at progressively lower peptide-vesicle ratios until the vesicle scatter accounted for half of the background signal at 300 nm. Binding is determined by spectral blue shifts and large intensity changes in

the presence of SUV. All peptides were first tested with neutral POPC SUVs, and if no binding was observed, 19:1 POPC:POPG SUVs were used. The fluorescence excitation wavelength was 280 nm, (290 nm for Gly3P-11 to avoid tyrosine excitation). Emission spectra were obtained from 300 to 400 nm (290 – 350 nm for TetA-4 tyrosine emission).

I also tested the temperature-dependent binding of the peptides, to see how suitable they would be for thermal titration or temperature jump studies. Temperature titrations were performed from 5 to 72 °C in the case of CCOIII-4 and 2-90 °C in the case of G3PT-4. Dry nitrogen was used to purge the sample chamber and the temperature probe was placed directly in the buffer blank cuvette. Peptide concentrations were between 10 and 30 µM and vesicle (not individual lipid) concentrations were between 0.15 and 0.4 µM.

### **Singular value decomposition (SVD).**

SVD was used to analyze fluorescence spectra obtained from thermal titrations. SVD is a linear algebra technique that decomposes an  $m \times n$  singular data matrix  $A$  into a column orthonormal  $m \times n$  matrix  $U$ , a column orthonormal  $n \times n$  basis function matrix  $V$ , and a diagonal  $n \times n$  singular value matrix  $W$

$$A = U \cdot W \cdot V^T \quad (1)$$

In the case of dispersed fluorescence melts,  $A$  contains fluorescence spectra at different temperatures. The resulting  $V$  contains basis function spectra,  $W$  tells us how important each basis function is, and  $U$  gives the temperature dependence of each basis function.

The SVD technique can reveal many features not obvious in raw spectral data, such as two state behavior, spectral shifts, broadening and cooperativity. Two state behavior can be inferred when bases with shifted fluorescence maxima, exhibit a temperature dependent transition with the same midpoint,  $T_m$ . This is because the presence of populated intermediates would not be likely to produce the same dependence at two different wavelengths. Spectral shifts are visible with SVD analysis by the presence of basis spectra with different peak wavelengths, which exhibit different temperature dependence, so that intensity will be shifted from one region of the spectrum to another as the temperature is changed. Broadening is observed when a basis that crosses the independent axis twice also has a temperature trend that is not flat. A cooperative transition is observed when the right singular vectors exhibit a sigmoidal shape as a function of temperature. This will occur whenever the system has sufficiently many degrees of freedom, so that a linear change in free energy is produced as the temperature is varied. The sigmoidal shape is then produced in any signal (such as CD, fluorescence lifetime, integrated fluorescence, etc.), which is a function of the fraction of products or reactants. This is just a result of the exponential relationship between the equilibrium constant and free energy.

In general, for thermal titrations of proteins, the first basis function shows the temperature-averaged spectrum, and its trend will track the overall intensity change with temperature. Higher basis functions reveal spectral shifts, broadening or other more subtle features of the data, though there are generally no more than two additional features. In this study I plot normalized basis functions and include the singular value in the temperature trend to recover the total signal.

## Training parameters .

With relatively small input data sets, learning algorithms based on neural networks and linear programming are superior to least squares methods.<sup>11</sup> The goal of such algorithms is to separate objects into “good” (soluble peptides) and “bad” (insoluble peptides) in a multidimensional parameter space. As input for the learning algorithm, easily calculable sequence-based parameters must be defined for the peptides. The ultimate goal is to reduce the set of parameters to the lowest-dimensional linear combination that can separate peptides according to whether or not they are soluble at micromolar levels.

*Hydrophobicity* is calculated as the sum of transfer free energies from water to a low dielectric medium for all residues in the peptide. Two hydrophobicity scales were tested, representing the free energies of transfer of poly-leucine pentapeptides containing the test residue from water into n-octanol<sup>12</sup> as well as retention times in reversed phase HPLC using a trifluoroethanol/water gradient.<sup>13</sup>

*Amphipathicity* is calculated as the magnitude of the hydrophobic moment for the entire length of the peptide, as described by Eisenberg et al.<sup>14</sup> *Chargepoints* is calculated by summing up the number of charged residues positioned 3 or less residues away from the N or C terminus and subtracting the number of charged residues between position 4 and n-4, where n is the peptide length. The maximum hydrophobic stretch length, *StretchMax*, is calculated as the maximum number of consecutive hydrophobic residues in the peptide. The maximum hydrophobicity of any stretch, *HPMax*, is calculated by summing the hydrophobicity values for each stretch of consecutive hydrophobic residues, and then taking the most hydrophobic value.

Modified versions of the hydrophobicity scales were tested, in which alanine, proline and threonine were added in all possible combinations to the core hydrophobic residue list that includes cysteine, isoleucine, leucine, methionine, valine, phenylalanine, tyrosine and tryptophan. I also tested scales in which the alanine hydrophobicity is taken to be zero, since some of the initial training set peptides had unusually large numbers of alanine residues, which might skew results. With these variations in scales I tested 6 distinct *StretchMax*, 18 *HPMax*, 4 Hydrophobicity, 4 Amphipathicity and 1 Chargepoints parameter. Only parameters using the same hydrophobicity scale and set of hydrophobic residues were combined when looking for solutions in multi-dimensional space using the learning algorithm described below.

### **Learning algorithm.**

The algorithm optimizes a plane<sup>15</sup> separating “good” and “bad” peptides, and assigns peptides a positive or negative score based on distance to the plane and the side they are on. First a set of  $M$  training peptide sequences is split into two sets, soluble and insoluble, to train the algorithm. Next, an  $N$ -dimensional parameter space is chosen, and each peptide is assigned a vector  $\mathbf{X}$  in parameter space, e.g.  $\mathbf{X}_i = \{HPMax_i, Amphipathicity_i, StretchMax_i\}$  for peptide “i” in a 3-D parameter space. A random separator hyperplane with normal vector  $\mathbf{W}$  is chosen to initiate the algorithm. In each training iteration, a Boolean error is computed for a randomly chosen peptide “j” and then used to rotate the separator plane to include the misplaced point. To accomplish this, the dot product  $\mathbf{X}_j \cdot \mathbf{W}$  is taken. If the chosen training peptide is an element of the soluble set  $P$ , but the dot product is negative, or *vice-versa*, then the peptide was

misclassified. The solution is updated by computing  $\mathbf{W}_{\text{new}} = \mathbf{W} \mp \mathbf{X}_j$  and rotating the hyperplane to match  $\mathbf{W}_{\text{new}}$ . Since there is typically not a unique solution and it is not necessary that the optimal solution will pass through the centroid of the data sets, the separator plane is also translated on a grid. The optimal solution is taken as the one which minimizes the sum of the peptide distances  $d'$  in each set to an auxiliary plane running through that set's centroid and parallel to the separator plane, while maximizing the distance  $d$  of all peptides on each side of the separator plane (fig. 2.1). Because many similarly good solutions may be found with slightly different hydrophobic residue lists and hydrophobicity scales, I applied a robustness criterion: the optimal hydrophobicity scale and hydrophobic residue list were the ones producing the most solutions with the smallest number of parameters.

### **Database mining.**

The final criterion was used to mine the database of all known transmembrane helices obtained from the online MPtopo database.<sup>9</sup> This database is separated into three categories. The 3D\_Helix category is for sequences from proteins with known X-ray crystallography or NMR structures, the 1D\_Helix is for sequences which are known to be trans-membrane due to indirect evidence obtained from gene fusion, proteolytic degradation, and amino acid deletion experiments, and the 3D\_Other category (not used in this study) is for all sequences from monotopic or beta-barrel trans-membrane proteins with known X-ray crystal or NMR structures. The best plane separating all insoluble from soluble training set peptides was used to evaluate all the sequences in both the 3D\_Helix and 1D\_Helix databases.



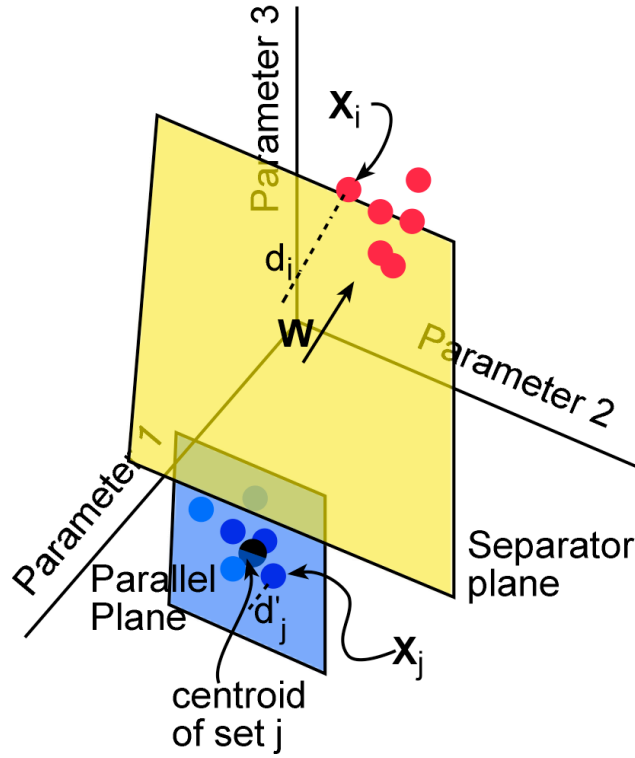
## 2.3 Results

### Training set characterization

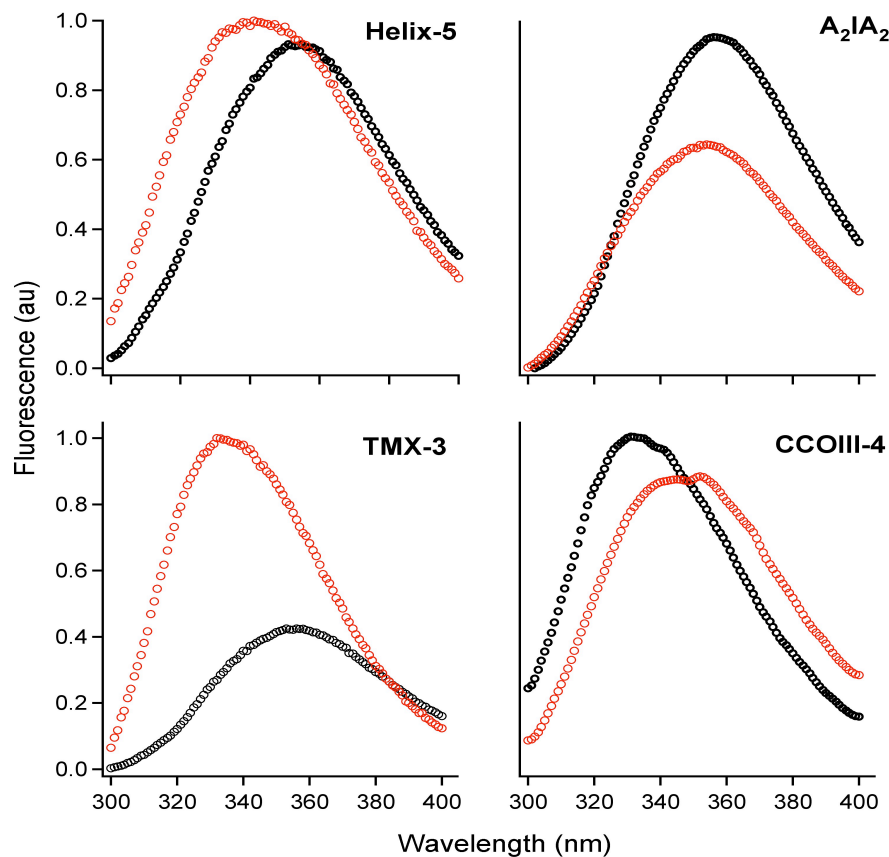
Table 2.1 lists the initial set of peptides that were used to train the learning algorithm, along with their measured solubility limits. Vpu and TMX-1 solubilities were below the detectable limit and Msba-1, PRC-1 and SRII-2 were all too hydrophobic to survive aqueous purification after synthesis and are assumed to be soluble to less than 1  $\mu\text{M}$ .

Figure 2.2 shows fluorescence spectra at 5°C for all soluble peptides from the training set with and without SUV. Helix-5 did not bind to neutral SUVs, but bound to anionic vesicles, producing a 13 nm blue shift at 5°C and a peptide-vesicle ratio of 72. CCOIII-4, TMX-3 and A<sub>2</sub>IA<sub>2</sub> bound strongly to both neutral and anionic SUVs. They produced spectral blue shifts of 14 nm, 22 nm and 5 nm, at peptide to anionic vesicle ratios of 100, 102, and 36, respectively. A<sub>2</sub>IA<sub>2</sub> and TMX-3 fluorescence intensity changed significantly upon binding. TMX-3 fluorescence more than doubles after binding, most likely due to less tryptophan quenching from the peptide's two histidine residues, which must be uncharged in order for the peptide to bind. A 33% quenching is observed upon A<sub>2</sub>IA<sub>2</sub> binding, though the reason for this is unclear.

**Figure 2.1.** The optimization algorithm rotates and translates the yellow plane to maximize the distances  $d_i$  from the plane of peptides “i” characterized by parameter vectors  $\mathbf{X}_i$ , while minimizing the distances  $d_j'$  to similar planes (one of two shown in blue) going through the centroid of each set to be classified.



**Figure 2.2.** Fluorescence spectra of training set peptides with (red circles) and without (black circles) anionic SUV. Spectra are taken at 5°C where binding is expected to be maximal. Blue spectral shifts are observed for all peptides in the presence of SUV. In addition, TMX-3 fluorescence is doubled, while A<sub>2</sub>IA<sub>2</sub> binding is quenched upon binding.



**Table 2.1.** Names and sources of training set peptides along with their amino acid sequences and measured solubilities. Changes to the native/designed sequences are shown in bold above substituted residues. Ac- = acetylation, -NH<sub>2</sub> = amidation.

Name/Source	Sequence	Solubility
Vpu <sub>8-29+</sub> HIV-1 Protease <sup>16</sup>	<b>KKG</b> IVALVV <del>IIII</del> IAIVVWSIV <del>II</del> E <b>GKKK</b>	< 1 μM
TMX-1 <i>de novo</i> <sup>17</sup>	Ac-WNALAAVAAALAAVAAALAAVAASKSKSKSK-NH <sub>2</sub>	< 1 μM
Helix-5 <i>de novo</i> <sup>6</sup>	Ac-NELKKKLELCKAKWLEAKKKLEALK-NH <sub>2</sub>	> 13 mM
TMX-3 <i>de novo</i> <sup>7</sup>	GGWAALAAHLAPALAAALAHALASRSRSR-NH <sub>2</sub>	14 μM
A <sub>2</sub> IA <sub>2</sub> <i>de novo</i> <sup>18</sup>	KKAAAI <del>AAAA</del> IAAWAAIAAKKKK-NH <sub>2</sub>	> 13 mM
Msba-1 Transport ATP-binding protein Msba <sup>19</sup>	Ac- <sup>H</sup> <sub>K</sub> AGLVVSTIALVINAAADT <sup>W</sup> <sub>Y</sub> MISLLK <sup>S</sup> <sub>PL</sub>	< 1 μM
PRC-1 Photosynthetic Reaction Center <sup>20</sup>	<b>RS</b> <sub>G</sub> AGIAAFAFGSTAILI <sup>A</sup> <sub>I</sub> LFN <sup>W</sup> <sub>M</sub> AA <sub>E</sub> VH-NH <sub>2</sub>	< 1 μM
CCOIII-4 Cytochrome C Oxidase Polypeptide III <sup>21</sup>	Ac- <sup>A</sup> <sub>L</sub> PLINTLI <sup>A</sup> <sub>L</sub> <sup>A</sup> <sub>L</sub> SGVAVTWAHHAFVKS	16 μM
SRII-2 Sensory Rhodopsin II <sup>22</sup>	<sup>S</sup> <sub>E</sub> RR <sup>H</sup> <sub>Y</sub> YVTLVGISGIAAVA <sup>W</sup> <sub>Y</sub> VVMAL <sup>H</sup> -NH <sub>2</sub>	< 1 μM

## Learning algorithm solutions

I were able to obtain several complete separations of the training set with as few as two parameters. Only 6 of these solutions are from the n-Octanol hydrophobicity scale, while the other 26 were found using the trifluoroethanol hydrophobicity scale, which I retain as more robust. I also retain the hydrophobic residue lists ACFILMVYW and APCFILMVYW as more robust (each produced 10 solutions) over the ATPCFILMVYW list (6 solutions). With that scale and list, no additional solutions were obtained when setting the alanine hydrophobicity value to zero. Thus I retained the trifluoroethanol scale and the hydrophobic residue list ACFILMVYW with alanine hydrophobicity at its experimentally determined value of 4.1 as our final choice.

The optimal two parameter solutions were the *HPMax – Amphipathicity* and *HPMax – Hydrophobicity*, with the former performing slightly better. Figure 2.3A shows the optimal *HPMax - Amphipathicity* solution. The dashed yellow line is the optimal separator line, and the yellow arrow indicates the single rotated parameter *D* that scores all soluble (red circles) and insoluble (blue circles) peptides in 1 dimension according to  $D > 0$  soluble,  $D < 0$  insoluble:

$$D = -\text{Amphipathicity} - 1.13 * \text{HPMax} + 354. \quad (1)$$

Though a perfect solution was obtained with just these two parameters, adding *StretchMax* in the 3<sup>rd</sup> dimension improves the final distance score by a factor of 2, more than any other added parameter. Figure 2.3B shows the 3-dimensional separator plot.

The optimal solution in this space is shown by the yellow plane, which predicts solubility based on the following inequality:

$$D_2 = -Amphipathicity - 0.845*HPMax - 0.124*StretchMax + 338 > 0. \quad (2)$$

It is clear from both plots in figure 2.3 that all peptides except for A<sub>2</sub>IA<sub>2</sub> can be separated using the HPMax parameter alone.

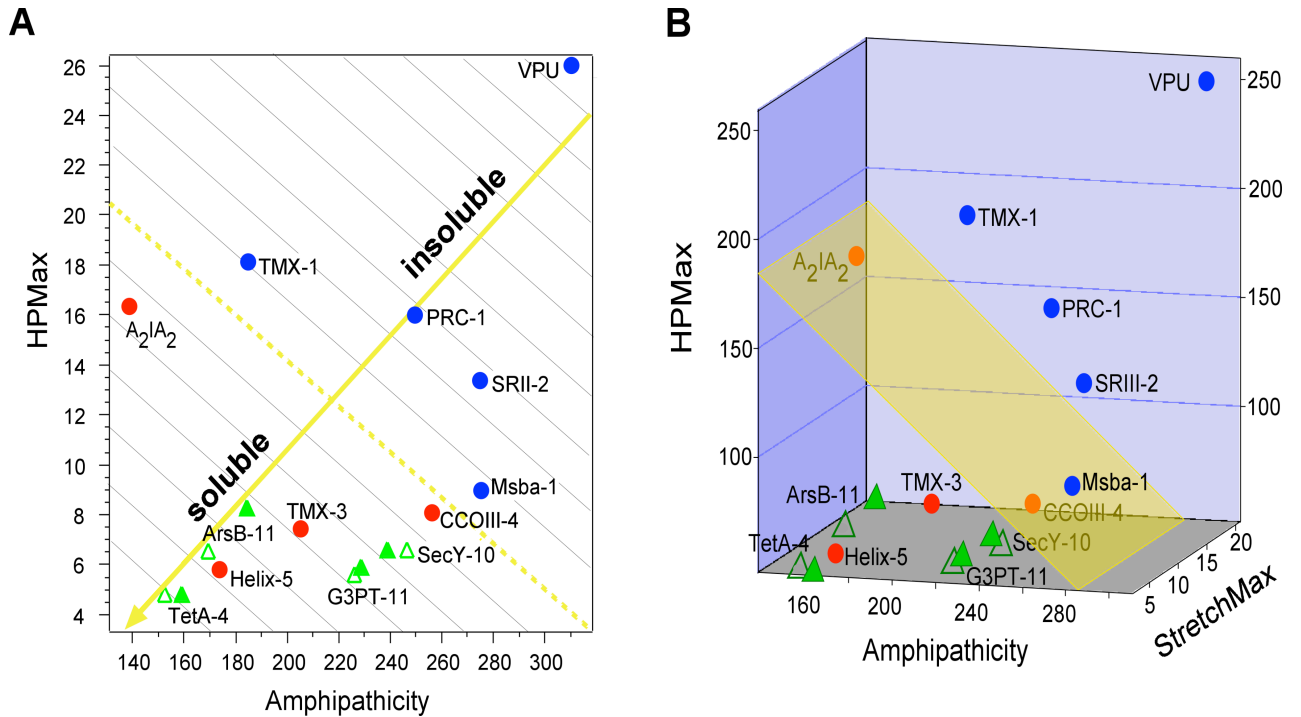
### **Selection of the validation set**

Next I needed to test the predictivity of eqs. (1) and (2). Figure 2.4 shows the distribution of  $D$  from the 1D\_Helix and 3D\_Helix databases. All sequences which satisfy eq. (2) also satisfy eq. (1), this includes a total of 92 out of 441 sequences in the 3D\_Helix database and 80 out of 261 sequences in the 1D\_Helix database.

I applied additional criteria to narrow the choice. Helices shorter than 21 residues were eliminated, leaving 78 candidates from both databases that can span the POPC/POPG membrane. Their  $D$  values ranged from 0.5 to 271 with lengths up to 32 residues. For the soluble training set,  $D$  values were 115, 65, 31, and 7, for Helix-5, TMX-3, A<sub>2</sub>IA<sub>2</sub>, and CCOIII-4, respectively. Since only A<sub>2</sub>IA<sub>2</sub> and Helix-5 were highly soluble, only peptides with scores  $D > 31$  were retained. I also eliminated peptides lacking a tyrosine or tryptophan fluorophore necessary for kinetics measurements. With these three criteria in place, 17 sequences remained from the 1D\_Helix database. The two with the largest  $D$  values, TetA-4 and ArsB-11, were chosen for synthesis.

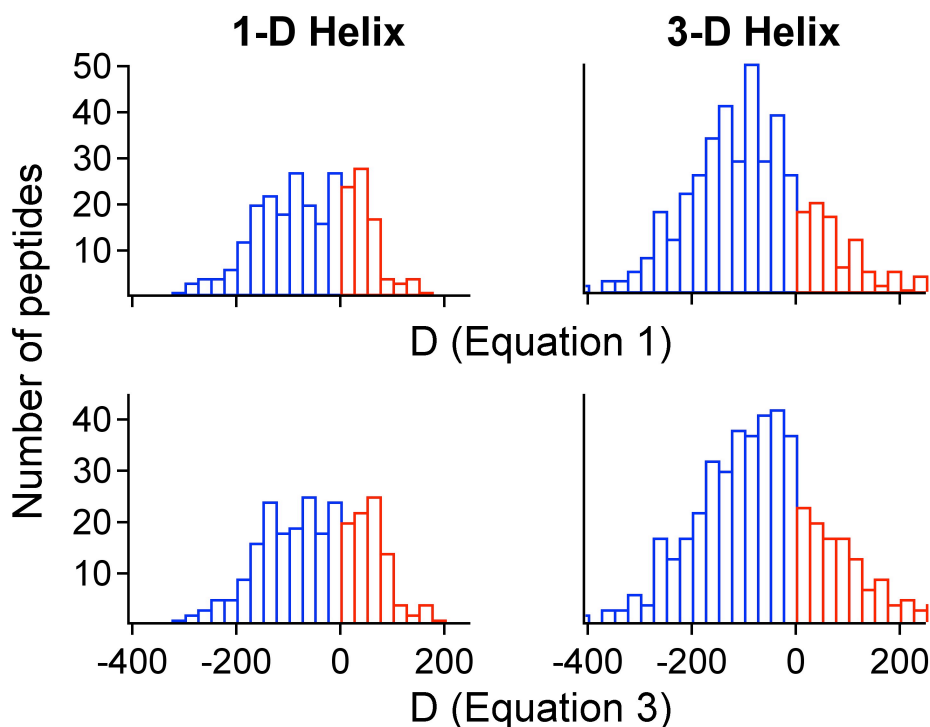
The 7 sequences remaining from the 3D\_Helix database after the above selection process were also examined structurally. Evidence of salt bridges or other significant protein-protein contacts which might be required for a helix to be stable inside the membrane, led to its exclusion. 3 of the 7 peptides contained one or more salt bridges and two of them were in the center of helical bundles where many protein-protein contacts were probable. Four helices lacked any salt bridges and appeared to require few contacts with the rest of the protein. The peptide that looked best from this visual inspection, G3PT-11, and the one that looked the worst, SecY-10, were chosen for synthesis to determine whether such visual inspection is worthwhile in the future. These peptides are listed in table 2.2 along with the two 1D\_Helix sequences chosen. Before synthesis, minor changes were made to the sequences to improve solubility or fluorescence detection. These are shown in bold in table 2.2, above the replaced residues. Parameters for these sequences are also plotted in figure 2.3, where open green triangles represent the wild type sequences, and closed green triangles represent the actual synthesized sequences.

**Figure 2.3** **A.** Optimal solution using *Amphipathicity* and *HPMax* calculated using the trifluoroethanol scale, the hydrophobic residue list ACFILMVYW and with the alanine hydrophobicity value unchanged. The yellow arrow indicates the rotated coordinate which correctly classifies all soluble (red circles) and insoluble (blue circles) training set peptides in just 1-dimension. Distance scores are indicated by black dotted lines and calculable using equation 1. **B.** Adding *StretchMax* as the 3<sup>rd</sup> parameter doubles the optimization score. The separation plane is shown in yellow. Both wild type (open triangles) and synthesized (closed triangles) are also shown in green in both plots.





**Figure 2.4** Distribution of scores, calculated according to the 2-dimensional solutions given in equations 1 and 3 for sequences in the 3D\_Helix and 1D\_Helix databases. Peptides with positive scores, shown in red, are expected to be soluble while peptides with zero or negative scores, shown in blue, are expected to be insoluble.



### Characterization of the validation set

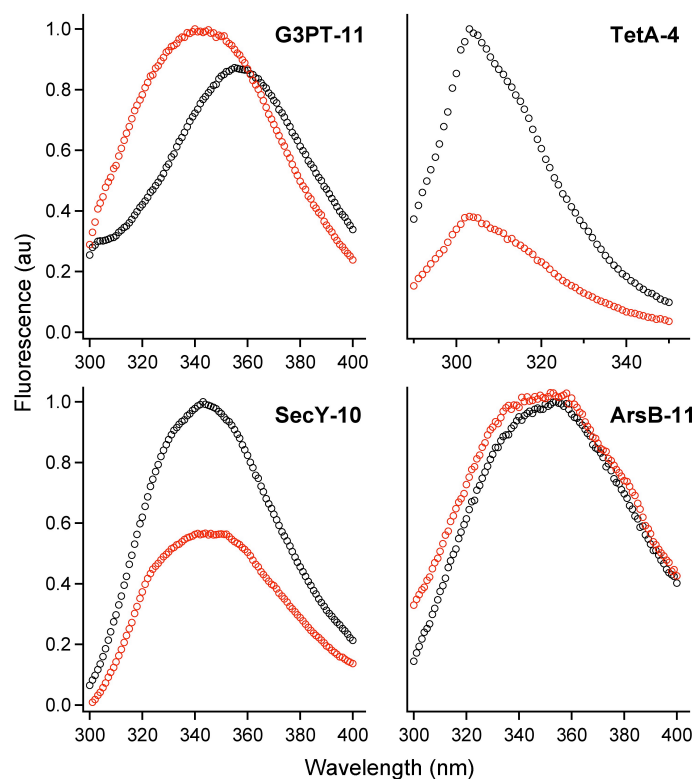
All four of the validation set peptides proved to be soluble and the observed solubility limits are listed in table 2.2. No significant correlation was observed, however, between distance scores calculated according to either equation (1) or (2) and observed solubility. Figure 2.5 shows fluorescence spectra at 5°C for all peptides in the validation set before and after addition of anionic SUV. A large blue shift of 13 nm was observed for G3PT-11 at a peptide-vesicle ratio of 73. Strong fluorescence quenching was observed for TetA-4 (62%) and SecY-10 (44%) upon the addition of anionic vesicle at

peptide-vesicle ratios of 250 and 9, respectively. For the final peptide, ArsB-11, no evidence of binding was observed for peptide-vesicle ratios as low as 12 in either neutral or anionic SUVs. This peptide came from the 1D\_helix database, where no structural check could be made. Thus three of the four peptides have fluorescence changes that can be used to monitor binding equilibria or kinetics, and both peptides from the 3D\_Helix database were water-soluble and showed fluorescence shifts indicative of binding.

**Table 2.2** Sequences chosen for validation of the learning algorithm solution. Mutations to the native sequences are shown in bold above substituted residues along with their measured solubilities.

Name	Sequence	Solubility
ArsB-11, Arsenical Pump Membrane Protein <sup>23</sup>	Ac-STATGVKEAMI <sup><b>W</b></sup> <sub>Y</sub> ANVIGCDL	25 $\mu$ M
SecY-10, Preprotein Translocase Secy Subunit <sup>24</sup>	Ac-GTGVLLTVSIVYR <sup><b>W</b></sup> <sub>M</sub> YEQLREKVSE <sup><b>G</b></sup> <sub>L</sub> -NH <sub>2</sub>	25 $\mu$ M
G3PT-11, Glycerol 3 Phosphate Transporter <sup>25</sup>	Ac-AAGTAAGFTGL <sup><b>W</b></sup> <sub>F</sub> GYLGGSVAASAIVGYTVD-NH <sub>2</sub>	2 mM
TetA-4, Tetracycline Resistance Protein <sup>26</sup>	Ac- <sup><b>GT</b></sup> IVAGITGATGAVAGAYIADIT-NH <sub>2</sub>	600 $\mu$ M

**Figure 2.5** Fluorescence spectra of validation set peptides with (red circles) and without (black circles) anionic SUV. Spectra are taken at 5°C where binding is expected to be maximal. Blue spectral shifts are observed only for G3PT-11, though significant quenching is observed for both TetA-4 and SecY-10 in the presence of SUV. No spectral or intensity shifts are observed for ArsB-11 with SUV at peptide-vesicle ratios as low as 12.



## 2.4 Discussion

I have shown that a simple scalar parameter  $D$  can be used to select soluble peptides suitable for peptide-membrane binding studies from membrane peptide databases. Out of a combined database of 702 helices, I applied  $D$  together with additional length and structural criteria to select 4 sequences. Success was achieved for

solubility of all 4 helices, and membrane binding was observed for 3 out of 4 peptides over the range of peptide-vesicle ratios feasible for fluorescence measurements.

If I incorporate the validation set peptides into the training set of our learning algorithm, I get a slightly different and updated condition for solubility analogous to eq. (1):

$$D = -Amphipathicity - 1.21HPMax + 378 > 0, \quad (3)$$

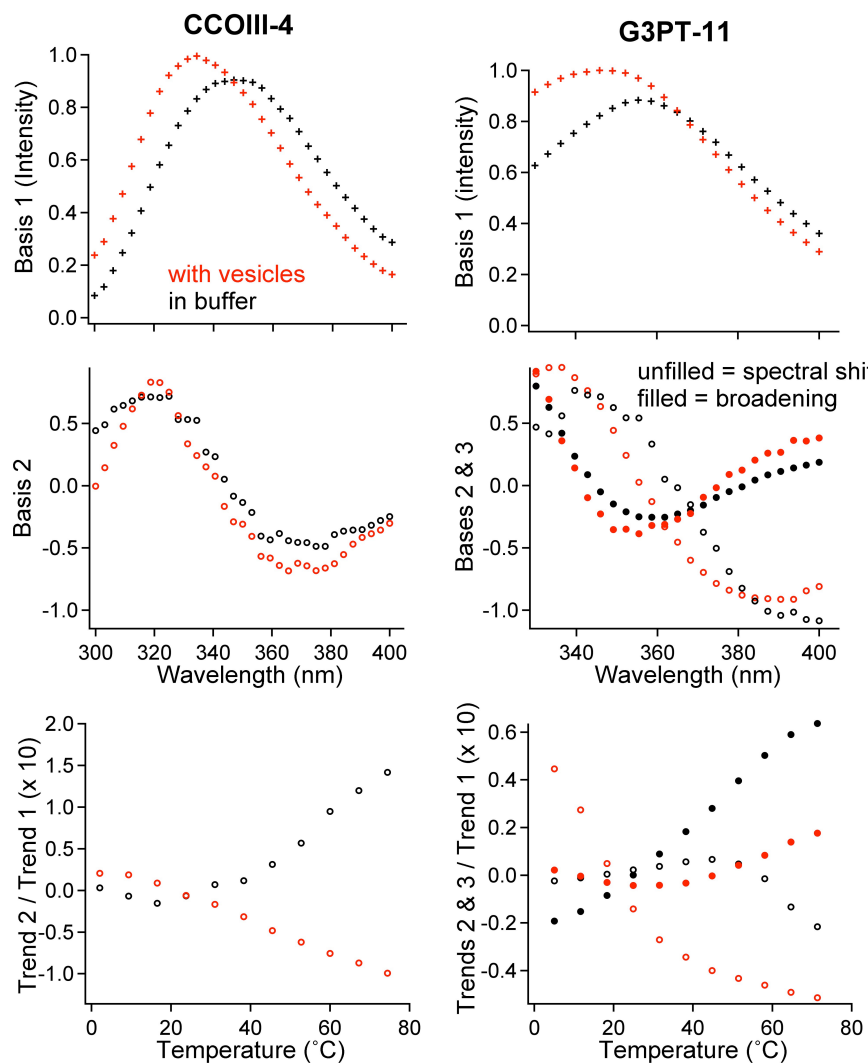
This updated solution can again be applied to the 1D\_Helix and 3D\_Helix databases and figure 2.4 shows the distribution of  $D$  values for both databases using this new solution. 92 and 103 peptides from the 1D\_Helix and 3D\_Helix databases respectively satisfy eq. (3). Table 2.3 shows that this new  $D$  correctly predicts 3 out of 4 peptides from the literature (not including TMX-3 that was used for training) known to have membrane binding constants near 1.

Application of this algorithm has already significantly increased the number of helical membrane peptides accessible to binding and folding experiments. As an example, figure 2.6 shows the singular value decomposition (SVD) analysis of temperature titrations for one peptide from the training set, CCOIII-4, and one peptide from the validation set, G3PT-11. Both are novel peptides based on natural trans-membrane helices with known 3-dimensional structures. The top panel shows that CCOIII-4 has a significant blue shift, while G3PT-11 both blue shifts and broadens upon vesicle binding. This is confirmed by the bottom panels in figure 2.6, which show large temperature trends obtained for the SVD basis functions in the middle panel. The trends

are very different in the presence of vesicles, allowing easy identification of binding equilibrium changes with temperature. The spectral shifts in vesicles occur due to burial of tryptophan in the hydrophobic vesicle interior; the generally opposite trends observed in pure buffer are due to temperature-dependent structural changes or dimerization of the peptides in solution.

I conclude by making additional predictions for soluble membrane-binding peptides. Table 2.4 shows 4 new peptides from the 3D\_Helix database that satisfy all the criteria used to select the validation set using the updated D in eq. 3. These peptides were ones that passed the visual inspection described in the results section above. This was deemed worthwhile since the aqueous solubility and fluorescence shift upon membrane binding observed for the peptide which passed such a visual inspection, G3PT-11, was significantly better than SecY-10, which did not pass. I believe that these additional sequences, perhaps with minor modification to improve fluorescence detection, are likely to yield useful peptides for binding kinetics and thermodynamics experiments.

**Figure 2.6.** SVD analysis of thermal titrations are shown for CCOIII-4 and G3PT-11. **Top:** First SVD basis functions correspond to average fluorescence spectra with (red) and without (black) anionic SUV. **Middle:** Basis functions 2 and 3 represent different physical features such as spectral shifts (open circles) and spectral broadening (filled circles). **Bottom:** Temperature trends of basis functions 2 and 3 are ratioed by trend 1 to remove any intensity changes which do not differ in the presence and absence of SUV. This analysis reveals spectral shifts in both peptides which go in opposite directions depending on the presence or absence SUV.



**Table 2.3** *D* scores calculated using eq. 3 are shown for peptides studied by other groups that exhibit near-unity equilibrium constants between water and membranes. Phenylalanine parameters were used for the modified residue in Magainin-2-P1.

Name	Sequence	<i>D</i>
Mastoparan-X <sup>3</sup>	INWKGIAAMAKKLL	175
Melittin <sup>1</sup>	GIGAVLKVLTTGLPALISWIKRKRQQ	76
Magainin-2-P1 <sup>4</sup>	Phe <sub>CN</sub> -GIGKFLHSACK-WGKAFVGQIMNS	96
Anti- $\alpha_{IIB}$ <sup>5</sup>	KKAYVMLLPFFIGLLLGLIFGGAFWGPARGHLKK	-156

**Table 2.4** List of additional sequences from the 3D\_Helix database which satisfy all criteria used in selecting the validation set for this study. The distance scores are calculated using equation 3.

Source (PDB ID)	Peptide Sequence	<i>D</i>
Glycerol-3-phosphate transporter (1PW4)	DLGFALSGISIAYGFSKFIMGSVSD	49
ADP, ATP carrier protein heart isoform T1 (1OKC)	FNVSQVQGIHYYRAAYFGVYDTAKG	66
Cytochrome c oxidase polypeptide III (1OCC)	VPLLNTSVLLASGVSITWAHHSLM	44
Lactose Permease (1PV6)	GEQGTRVFGYVTTMGELNASIM	131

## 2.5 References

1. Constantinescu, I. & Lafleur, M. (2004). Influence of the lipid composition on the kinetics of concerted insertion and folding of melittin in bilayers. *Biochimica et Biophysica Acta* **1667**, 26-37.
2. Tang, J. & Gai, F. (2008). Dissecting the Membrane Binding and Insertion Kinetics of a pHLIP Peptide. *Biochemistry* **47**, 8250-8252.
3. Tang, J., Signarvic, R. S., DeGrado, W. F. & Gai, F. (2007). Role of Helix Nucleation in the Kinetics of Binding of Mastoparan X to Phospholipid Bilayers. *Biochemistry* **46**, 13856-13863.
4. Tucker, M. J., Tang, J. & Gai, F. (2006). Probing the Kinetics of Membrane-Mediated Helix Folding. *Journal of Physical Chemistry B* **110**, 8105-8109.
5. Tang, J., Yin, H., Qiu, J., Tucker, M. J., DeGrado, W. F. & Gai, F. (2009). Using Two Fluorescent Probes to Dissect the Binding, Insertion, and Dimerization Kinetics of a Model Membrane Peptide. *Journal of the American Chemical Society* **131**, 3816-3817.
6. Russel, C. J., King, D. S., Thorgeirsson, T. E. & Shin, Y. K. (1998). De novo design of a peptide which partitions between water and phospholipid bilayers as a monomeric alpha-helix. *Protein Engineering* **11**, 539-547.
7. Ladokhin, A. S. & White, S. H. (2004). Interfacial Folding and Membrane Insertion of a Designed Helical Peptide. *Biochemistry* **43**, 5782-5791.
8. Reshetnyak, Y. K., Segala, M., Andreev, O. A. & Engelman, D. M. (2007). A Monomeric Membrane Peptide that Lives in Three Worlds: In Solution,



- Attached to, and Inserted across Lipid Bilayers. *Biophysical Journal* **93**, 2363-2372.
9. Jayasinghe, S., Hristova, K. & White, S. H. (2001). MPtopo: A database of membrane protein topology. *Protein Science* **10**, 455-458.
  10. Gill SC, v. H. P. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **182**, 319-326.
  11. Rojas, N. L., Kamtekar., S., Simons, C. T., McLean, J. E., Vogel, K. M., Spiro, T. G., Farid, R. S. & Hecht, M. H. (1997). De novo heme proteins from designed combinatorial libraries. *Protein. Sci.* **6**, 2512-2524.
  12. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109-5124.
  13. Chen, Y., Mant, C. T., Farmer, S. W., Hancock, R. E. W., Vasil, M. L. & Hodges, R. S. (2005). Rational Design of {alpha}-Helical Antimicrobial Peptides with Enhanced Activities and Specificity/Therapeutic Index. *Journal of Biological Chemistry* **280**, 12316-12329.
  14. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *Journal of Molecular Biology* **179**, 125-142.
  15. Rojas, R. (1996). *Neural Networks, A Systematic Introduction*, Springer-Verlag.
  16. Park, S. H., Mrse, A. A., Nevzorov, A. A., Mesleh, M. F., Oblatt-Montal, M., Montal, M. & Opella, S. J. (2003). Three-dimensional Structure of the Channel-

- forming Trans-membrane Domain of Virus Protein “u” (Vpu) from HIV-1. *Journal of Molecular Biology* **333**, 409-424.
17. Wimley, W. C. & White, S. H. (2000). Designing Transmembrane  $\alpha$ -Helices That Insert Spontaneously. *Biochemistry* **39**, 4432-4442.
  18. Liu, L.-P. & Deber, C. M. (1997). Anionic Phospholipids Modulate Peptide Insertion into Membranes. **36**, 5476-5482.
  19. Chang, G. (2003). Structure of MsbA from *Vibrio cholera*: A Multidrug Resistance ABC Transporter Homolog in a Closed Conformation. *Journal of Molecular Biology* **330**, 419-430.
  20. Deisenhofer, J., Epp, O., Sinning, I., Michel, H. (1995). Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodospseudomonas viridis*. *Journal of Molecular Biology* **246**, 429-457.
  21. Harrenga, A. & Michel, H. (1999). The Cytochrome C Oxidase from *Paracoccus denitrificans* Does not Change the Metal Center Ligation Upon Reduction. *Journal of Biological Chemistry* **274**, 33296.
  22. Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E. M., Pebay-Peyroula, E. & Navarro, J. (2001). X-Ray Structure of Sensory Rhodopsin II at 2.1 Å Resolution. *Proceedings of the National Academy of Science (PNAS)* **98**, 10131-10136.
  23. Tisa, L. S. & Rosen, B. P. (1990). Molecular characterization of an anion pump. The ArsB protein is the membrane anchor for the ArsA protein. *Journal of Biological Chemistry* **265**, 190-194.

24. van den Berg, B., Clemons Jr., W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44.
25. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. (2003). Structure and Mechanism of the Glycerol-3-Phosphate Transporter from Escherichia Coli. *Science* **301**, 616-620.
26. Waters S.H., R. P., Grinsted J., Altenbuchner J., Schmitt R. (1983). The tetracycline resistance determinants of RP1 and Tn1721: nucleotide sequence analysis. *Nucleic Acids Research* **11**, 6089-6105.

## Chapter 3

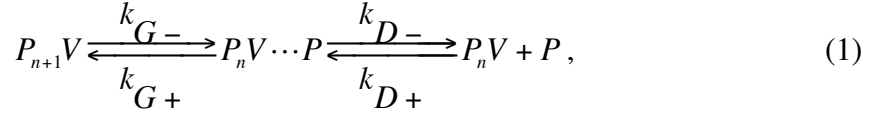
### Peptide-Membrane Association Dynamics

#### 3.1 Introduction

The association reaction of peptides with membranes is critical for formation of membrane channels in biomolecular recognition and signal transduction, and as a first step in membrane protein folding. Much progress has been made in studying the kinetics of helix association, folding, and insertion into membranes using stopped-flow and manual mixing experiments<sup>1; 2; 3; 4; 5</sup>. Experimental results support both the possibility of binding limited by secondary structure formation and followed by insertion, or secondary structure formation concomitant with insertion.

Association itself is already at least a two-step process, involving diffusive formation of an encounter complex between peptide and membrane surface, followed by geminate embedding of the peptide in the membrane surface, which provides a greater than 10 Å polar layer for accommodation due to surface dynamics of the membrane. Resolving the association reaction requires a peptide with pre-formed helical structure, a high concentration of surface binding sites, and kinetic methods faster than stopped-flow so that both steps can be brought to a similar time scale and observed simultaneously. Here we provide such results by studying the association-disassociation dynamics of the fluorescent helical peptide helix 5 with anionic SUV.

We propose the following scheme to describe peptide-membrane (dis)association,



where “G” indicates geminate recombination/dissociation of the strongly bound state, while “D” indicates diffusion to form/break the encounter complex.  $P_nV$  is the vesicle with  $n$  peptides already bound. Using this model, the rate equations are,

$$\begin{aligned} \frac{d[P]}{dt} &= \sum_{n=0} k_{D-}[P_nV \cdots P] - k_{D+}[P][P_nV], \quad \frac{d[V]}{dt} = -k_{D+}[V][P] + k_{D-}[V \cdots P] \\ \frac{d[P_nV]}{dt} &= +k_{G+}[P_{n-1}V \cdots P] - k_{G-}[P_nV], \quad n \geq 1 \\ \frac{d[P_nV \cdots P]}{dt} &= +k_{D+}[P][P_nV] - k_{D-}[P_nV \cdots P] - k_{G+}[P_nV \cdots P] + k_{G-}[P_{n+1}V], \quad n \geq 0 \end{aligned} \quad (2)$$

In our experiments, typical peptide vesicle ratios (PVR) are less than 100. Since the helix length is on the order of 5nm and there is about  $64\text{\AA}^2$  per lipid, there are approximately 400 independent peptide-binding sites for Helix 5 on a 50nm vesicle. “Independent binding site” means that the rate coefficients and change in spectroscopic signatures are the same whether a peptide binds to a vesicle without another peptide, or one that already has peptides bound.

If the binding sites are independent, then the signals from  $[P_nV]$  are the same for all  $n \geq 1$ . Only three contributions to the fluorescence exist, from  $P$ ,  $[P_nV]$  and  $[P_nV \cdots P]$ . Adding all of these together yields,

$$\begin{aligned} \frac{d[P]}{dt} &= k_{D-}[V \cdots P] - k_{D+}[P][V_0], \\ \frac{d[V_P]}{dt} &= +k_{G+}[V \cdots P] - k_{G-}[V_P], \quad n \geq 1 \\ \frac{d[V \cdots P]}{dt} &= +k_{D+}[P][V_0] - k_{D-}[V \cdots P] - k_{G+}[V \cdots P] + k_{G-}[V_P], \quad n \geq 0 \end{aligned} \quad (3)$$

where  $V_0$  is the total vesicle concentration,  $[V_p] = [PV] + [P_2V] + \dots$  is the concentration of vesicles with any number of peptides bound to it, and  $[V \cdots P] = [P_0V \cdots P] + [P_1V \cdots P] + \dots$  is the concentration of vesicle with one peptide associated and any other number of peptides already bound to it. Note that we are implicitly assuming that either  $V \cdots P$  is a higher energy state, or that the number of peptides per vesicle is sufficiently small, since we are not including cases like  $P \cdots V \cdots P$ , where more than one peptide only is associated with a vesicle at any given time. This assumption is likely valid at the low PVR values employed in our experiments, but this can be tested by measuring the kinetics at a range of concentrations.

The diffusional rate coefficient  $k_{D+}$  is given by,

$$k_{D+} (M^{-1}s^{-1}) \approx 10^3 4\pi(R_v + R_p)(D_v + D_p)A \approx 4\pi R_v D_p A \quad (4)$$

where the diffusion coefficient for the peptide is given by,

$$D_p A \approx \frac{RT}{6\pi\eta R_p} \approx \frac{2500J}{6\pi \cdot 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1} \cdot 10^{-9} \text{ m}} \quad (5)$$

assuming a viscosity of  $1 \text{ cP} = 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}$  and a peptide radius of  $1 \text{ nm}$ . For  $50 \text{ nm}$  radius vesicles, the bimolecular diffusion-limited rate becomes  $k_{D+} \approx 8 \cdot 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ . Thus at a vesicle concentration of  $V_0 = 0.5 \text{ } \mu\text{M}$ , the rate of peptide disappearance (irrespective for now of the rate of peptide appearance) is,

$$\frac{d[P]}{dt} = -k_{D+}[P][V_0] + k_{D-}[V \cdots P] \approx -\frac{[P]}{25 \mu s} + k_{D-}[V \cdots P] \quad (6)$$

This is expected to be about the fastest rate observed in the experiment at high concentration of vesicles (diffusion-limited collision of peptide and vesicle).

Making the steady-state approximation  $d[V\cdots P]/dt \approx 0$  (the loosely associated state is never populated), one obtains,

$$[V\cdots P] \approx \frac{k_{D+}[P][V_0] + k_{G-}[V_P]}{k_{D-} + k_{G+}} \quad (7)$$

From this, we can estimate the initial rate of disappearance of  $[V_P]$  in an upward temperature jump starting under strongly bound conditions,

$$\begin{aligned} \frac{d[V_P]}{dt} &= +k_{G+} \frac{k_{D+}[P][V_0] + k_{G-}[V_P]}{k_{D-} + k_{G+}} - k_{G-}[V_P] \Rightarrow \\ k_{eff} &\approx k_{G-} \left( 1 - \frac{k_{G+}}{k_{D-} + k_{G+}} \right) \approx k_{G-} \left( 1 - \frac{k_{G+}}{k_{D-}} \right) \approx k_{G-} \end{aligned} \quad (8)$$

In the second line, it was assumed that  $[P]$  is initially negligible (i.e. all peptide is bound), and that the diffusive dissociation rate is much faster than the geminate binding rate under unfavorable conditions right after the T-jump. In the extreme case, the measured rate is the geminate unbinding rate, and this should be true for large T-jumps beginning where the peptides are all bound, and ending at high temperatures where equilibrium favors complete dissociation. Under the same conditions, the steady-state approximation inserted into the first of eqs. (2) yields for the rate of appearance of peptide,

$$\begin{aligned} \frac{d[P]}{dt} &= +k_{D-} \frac{k_{D+}[V_0][P] + k_{G-}[V_P]}{k_{D-} + k_{G+}} - k_{D+}[V_0][P] \Rightarrow \\ k_{eff} &\approx \frac{k_{D-}k_{G-}}{k_{D-} + k_{G+}} \approx k_{G-} \end{aligned} \quad (9)$$

The second line again assumes that  $[P]$  is negligible just before the T-jump, and the extreme case is that  $k_{D-} \gg k_{G+}$  under dissociation (high temperature) conditions, leading to the simple relation in the second half of the line. Thus, in the extreme case,  $k_{G-}$  is the observed rate coefficient, not  $k_{D-}$ , however if the rates in eq. (1) are comparable, a full

solution of the linear differential equations is required. This will occur for T-jumps in the range where the transition is made, and will be useful, in combination with thermodynamic measurements, for separating out various rate coefficients.

## **3.2 Materials and Methods**

### **Buffers and sample preparation.**

As described in chapter 2, Helix 5 (Ac-NELKKKLELCKAKWLEAKKKLEA LK-NH<sub>2</sub>) is an amphipathic membrane-binding peptide, monomeric in solution up to at least 3mM which does not insert into the membrane after binding to its surface. Solid phase synthesis and HPLC purification was performed by Genscript, Inc. (Piscataway, NJ) and was verified > 95% by the UIUC School of Chemical Sciences Mass Spectrometry facility. Anionic SUV of ~50nm diameter are prepared and characterized as described in Chapter 1 using lipids purchased from Avanti Polar Lipids (Alabaster, AL). All measurements were performed in 50 mM sodium phosphate buffer. Protein – SUV mixtures are prepared using 2X vesicle and 10X peptide stock solutions. Final peptide-vesicle (not lipid) ratios are given as “PVR”. Using 64Å/lipid, there are approximately 12,000 outer leaflet lipids per 50nm SUV. All buffers are prepared using double distilled, deionized water and are filtered with a 0.2 micron nylon membrane. Temperature jump samples are prepared in buffer containing D<sub>2</sub>O at the indicated concentrations.



### **Circular Dichroism (CD) measurements.**

CD measurements were obtained using a JASCO (Japan Spectroscopic Co., Tokyo, Japan) Spectropolarimeter model J-715 fitted with a PTC temperature controller and a NesLab RTE 111 water bath. Solution spectra of Helix 5 were obtained at 6uM, solution melts at 9uM. All samples were measured in a stopper-top, 1 cm path quartz cuvette with a 4mm wide window and black walls (Starna catalog No. 29B/Q/10). Mineral oil and a teflon cap is used to prevent evaporation during temperature titrations. To obtain CD spectra, the samples were scanned in 1nm increments from 250 to 200nm and 10 scans were averaged. Spectra are taken from 5°C to 75°C at 5°C increments in both buffer and SUV. Temperature titrations, where the 222nm signal was averaged for 2 were also obtained over the range -4-88°C, scanning in 2°C increments. Raw data is expressed as left minus right circularly polarized light,  $\Delta\epsilon$ , in  $10^{-3}\text{deg}\cdot\text{M}^{-1}\cdot\text{cm}^{-1}$ . A more useful measure of CD which allows direct comparison to other experiments using different instruments, proteins, and concentrations, is the mean residue ellipticity,  $[\theta]$ , expressed in  $\text{deg}\cdot\text{cm}^2/\text{dmol}$ . The conversion from raw data to  $[\theta]$  can be performed using equation 1:

$$[\theta] = \frac{\Delta\epsilon}{10 \cdot N_r \cdot l \cdot [c]} \quad (10)$$

where  $N_r$  is the number of residues in the protein,  $l$  is the path length of the cell in cm, and  $[c]$  is the protein concentration in M.

### **Steady state dispersed fluorescence measurements.**

Fluorescence spectra were taken to test whether peptides were capable of binding to SUV. All spectra were obtained using a Varian (Palo Alto, CA) Cary Eclipse spectrofluorimeter as described previously in chapter 2.

### **Generation of the UV laser pulse train.**

A 5W, 532nm vertically polarized beam from a Spectra Physics Millennia Pro 10s diode laser is used to pump a MIRA Titanium:Sapphire laser. This laser is mode locked to produce 100 femtosecond, 840nm infrared pulses at a frequency of 71.4 MHz. The infrared beam is then sent to a home built frequency tripler composed of a  $\beta$ -barium borate (BBO) crystal for type I second harmonic generation, and a second BBO crystal for type I sum frequency generation of the second harmonic and the fundamental beams. The final 280nm tripled beam is separated from residual blue and fundamental light using a prism and is focused to 100 $\mu$ m at the sample cell by a 100mm focal length fused silica lens. This beam is shuttered in order to minimize photobleaching of the sample.

### **Generation of the nanosecond temperature jump.**

The 10ns heating pulses used to perturb the Helix 5 –SUV binding are generated by a Continuum Surelite Q-switched Nd:YAG laser operating at 10Hz. The 1064nm pulse is shifted to 1.9 $\mu$ m by stimulated Raman scattering in a 1.2m long H<sub>2</sub> tube at 360 PSI. A dichroic beam splitter reflects the remaining fundamental light onto a beam dump and passes the Raman shifted light. A prism is then used to separate out the anti-Stokes

and higher order stokes lines so that only the 1.9 $\mu$ m light is directed into the sample chamber. A 50:50 beam splitter is used to split the pulse into two arms of unequal length for counter propagation without the nonlinear effects due to temporal overlap of the pulses. Control of energy differences in the two arms can be controlled using a thin D<sub>2</sub>O cell to attenuate the more intense beam. D<sub>2</sub>O cells can also be placed upstream of the 50:50 beam splitter to lower the overall jump size. The heating pulse is focused approximately 2cm past the sample cell on each side so that the spot size at the sample cell is 1mm, about 10 times larger than the UV probe. For samples prepared in 50% D<sub>2</sub>O, the jump size is typically 15-18 °C.

#### **Time resolved dispersed fluorescence.**

As shown in figure 3.1, our setup allows for simultaneous detection of dispersed fluorescence and integrated fluorescence after the sample is probed with the continuous pulse train of 280 nm, 200 fs duration Ti:sapphire laser pulses (generated in the manner described above) spaced by 14 ns. The dispersed fluorescence is collected at 90° into an f/0.85 liquid light guide (oriel 77554) and focused onto the entrance of an f/3.9 Triax 180 monochromator by two fused silica lenses. The dispersion of approximately 3.5nm/mm at the exit slit is achieved by a holographic grating blazed at 330nm with 1200grooves/mm. This is imaged onto a 16 channel array PMT (Hamamatsu, R5900U-L16) with a 1mm pitch to give about 3.5nm bandwidth and a 56nm maximum detection range. In order to obtain sufficient signal to noise, approximately 14 pulses are integrated into each data point so that the time resolution of this setup is 200nsec. The instrument response after the jump is less than 1  $\mu$ sec as revealed by the 15-30°C jump of 38 $\mu$ M

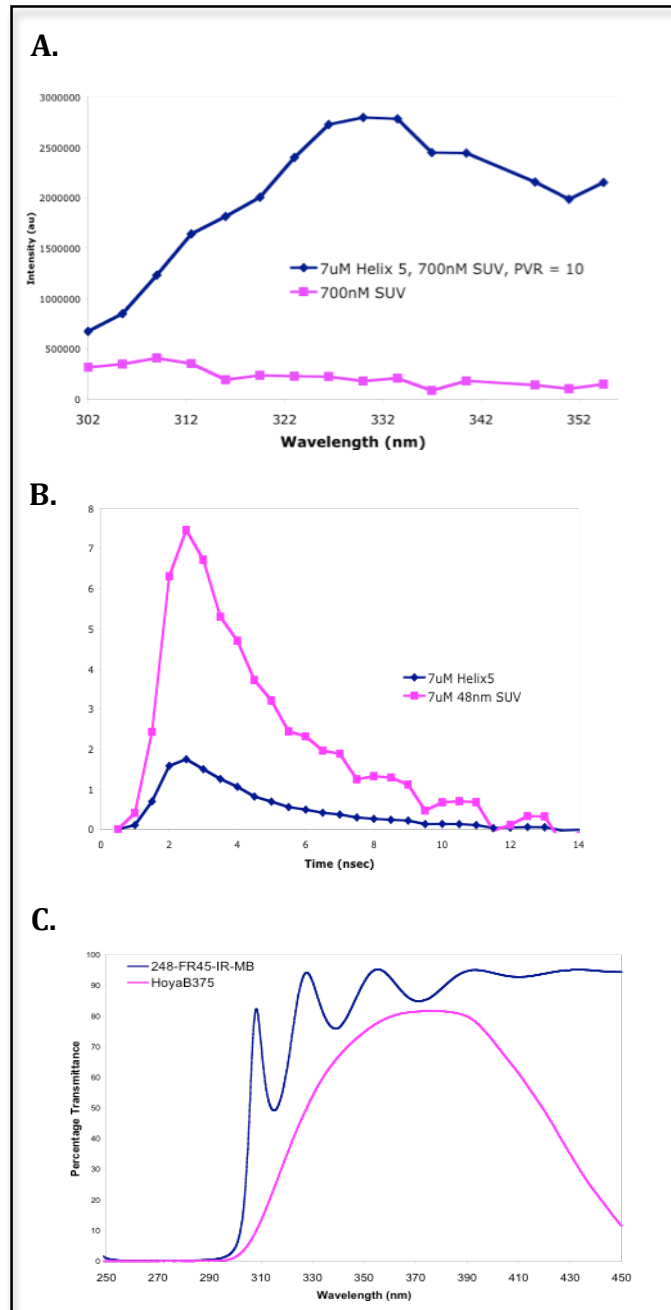
Tryptophan shown in figure 3.7. The spectra obtained with this setup are not smooth due to slightly different gains in each channel of the detector as shown in figure 3.2a but this is easily corrected by comparing spectra from fluorescence standards measured on this setup against that of a calibrated fluorometer. Fluorescence maxima obtained for Helix 5 in SUV at a PVR of 10 are 10 times greater than the scatter and fluorescence from SUV alone as shown in figure 3.2a. since most experiments are performed using a PVR near 30, SUV background signal is not a problem in these measurements.

### **Integrated fluorescence and fluorescence lifetime detection.**

The Helix 5 fluorescence, is measured using a single channel fast photomultiplier tube (PMT) with a rise time of 500 picoseconds. This signal is collected over 500 microseconds as the sample is probed every 14 nanoseconds with a 280 nm pulse. The analog data is digitized at 2 Gigaseimens per second with a 1 Ghz bandwidth to give 28 points per decay. Synchronization of the digitizer scope to the laser frequency is accomplished by external clocking from a photodiode which collects a portion of the light generated from a second harmonic generation crystal as described above. Figure 3.2b shows that scatter and background fluorescence from SUV are a problem, with a signal to background of less than 4 using a PVR as high as 145. Using an Acton mirror combined with a 375nm Hoya bandpass filter (see transmittance spectra in figure 3.2c) improves this signal/background ratio by a factor of 7.



**Figure 3.2** **A.** For a PVR=10, Helix 5 fluorescence in SUV (blue) is more than 10 times the scatter and background fluorescence from SUV alone (pink). **B.** The Helix 5 fluorescence in buffer (blue) is 4 times greater than the background from SUV when a PVR of 145 is used. This does not provide sufficient signal **C.** An Acton mirror (blue), used to block 99.9999% of the 280nm light, combined with a 375nm Hoya bandpass filter (pink) improves the signal to background by a factor of 7.



## Data Analysis Methods

### Singular value decomposition (SVD).

As described in chapter 2, SVD was used to generate temperature trends from the raw fluorescence spectra and spectrally resolved kinetics traces. We plot normalized basis functions and temperature trends which can be combined with the singular values listed to recover the total signal.

### Thermodynamic parameter fitting.

The equilibrium unfolding CD and fluorescence curves for Helix 5 in solution as well as the equilibrium unbinding-unfolding curves for Helix 5 in SUV are fit to a two state thermodynamic model in which the signal is represented as the fraction of unfolded or unbound peptide and the free energy is determined from a second order Taylor expansion about the transition temperature,  $T_o$ :

$$\Delta G = \Delta G_o + \Delta G_1(T-T_o) + \Delta G_2(T-T_o)^2 \quad (11)$$

The relationship between denatured/unbound fractions and temperature are obtained by expressing the signal as a linear combination of folded and unfolded signals:

$$S(T)_{b,f} = \frac{S(T)_{b,f} + S_{u,d} \cdot K_{eq}(T)}{1 + K_{eq}(T)} \quad (12)$$

where  $K_{eq}(T)$  is the equilibrium constant, defined by:

$$K_{eq}(T) = e^{-\Delta G(T)/RT} \quad (13)$$

$S(T)_{b,f}$  is the temperature dependence of the folded and/or bound state of the peptide at low temperature, and  $S(T)_{u,d}$  is the temperature dependence of the denatured and/or unbound state of the peptide at high temperature. These two baselines are assumed to be linear according to:

$$S(T)_{b,f} = S^{\circ}_{b,f} + m_{b,f} \cdot T \quad (14)$$

$$S(T)_{u,d} = S^{\circ}_{u,d} + m_{u,d} \cdot T \quad (15)$$

The success of this method relies heavily on the accuracy of (4) and (5), so that if any quadratic or higher order terms are present, this analysis will be unreliable.

### **Deconvolution of the fluorescence signal and determination of lifetimes.**

The signal collected after UV excitation of the protein sample will be a convolution of the tryptophan fluorescence decay and the instrument response function, IRF. The total signal is:

$$I(t) = \int R(t - t')S(t')dt' \quad (16)$$

where  $S(t)$  is the actual sample decay which is modeled by a single exponential:

$$S(t) = A_1 \exp(-t/\tau) \quad (17)$$



where  $\tau$  is the fluorescence lifetime, and  $R(t)$  is the detector time resolution, which we fit to a gaussian of width  $\sigma$ :

$$R(t) = A_1 \exp(-t^2/\sigma^2) \quad (18)$$

The measured signal,  $I(t)$ , is then fit to the convolution of these two

$$I(T) = \exp\left(\frac{\sigma^2}{2 \cdot \tau^2} - \frac{t}{\tau}\right) \cdot \left(1 - \operatorname{erf}\left(\frac{\sigma^2 - \tau \cdot t}{\sqrt{2 \cdot \sigma \cdot \tau}}\right)\right) \quad (19)$$

where  $\operatorname{erf}(x)$  is the error function such that  $\operatorname{erf}(0) = 0$  and  $\operatorname{erf}(\infty) = 1$ . In steady state measurements, all 500 microseconds, or about 36,000 transients are summed together before this fitting is performed.

### **Chi analysis to determine rates.**

Approximately 70 transients, or 1 microsecond worth of data is summed just before the temperature jump in order to obtain the low temperature profile,  $f_1$ , and about the same number are summed about 400 microseconds after the jump to obtain the high temperature profile,  $f_2$ . The total fluorescence at any time,  $t$ , after the temperature jump, will just be the sum of these two profiles weighted by their coefficients,  $\chi_1$  and  $\chi_2$ , respectively:

$$f(t) = \chi_1 f_1 + (1 - \chi_2) f_2 \quad (20)$$

A plot of  $\chi_1$  versus time can then be fit to a single exponential of the form:

$$\chi_1 = \alpha + A \exp(-t/k_{\text{obs1}}) \quad (21)$$

or a double exponential of the form:

$$\chi_1 = \alpha + A_1 \exp(-t/k_{\text{obs1}}) + A_2 \exp(-t/k_{\text{obs2}}) \quad (22)$$

where  $k_{\text{obs1}}$  and  $k_{\text{obs2}}$  are the observed rate constants for folding and membrane binding.

This same analysis can be performed on dispersed fluorescence data by fitting the above exponential form to either the raw decays observed in individual channels, or on the decay of the SVD basis functions.

### 3.3 Results

Circular dichroism thermal titrations and 20°C spectra for Helix 5 with and without SUV are shown in figure 3.3 B and A, respectively. Titrations in buffer reveal a smooth transition from a mean residue ellipticity of  $[\theta]_{222} = -25,000 \text{ deg cm}^2 \text{ dmol}^{-1}$  at 2 °C to -5,000 at 75 °C in buffer and -20,000  $\text{deg cm}^2 \text{ dmol}^{-1}$  to -12,000 in buffer. These high mean residue ellipticities are in keeping with the high helix propensity (AGADIR score of 49% at 25 °C, pH 7, and 0.1 ionic strength). The helical structure of this peptide is thus largely pre-formed when it interacts with the membrane, so helix formation should not substantially slow down the association process. Furthermore, the membrane clearly

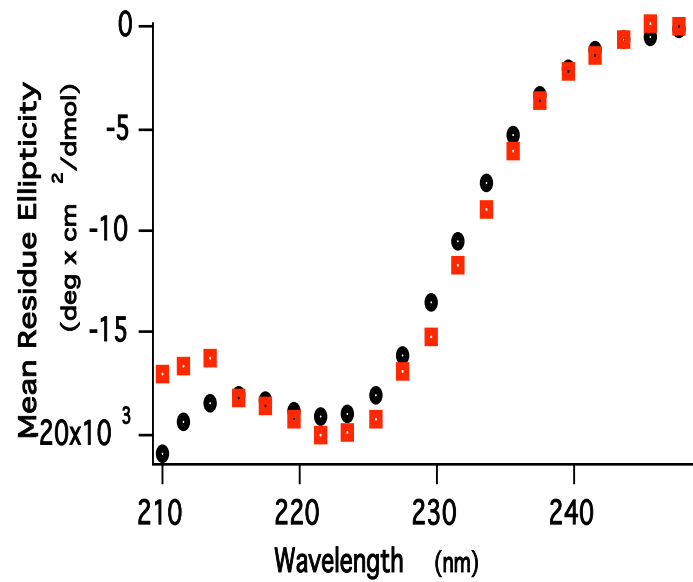
stabilizes the helical structure against thermally induced unfolding relative to the peptide in buffer.

Figure 3.4A shows Helix 5 fluorescence spectra at 5 and 75°C with and without SUV. The sample shown contains 19 mM peptide in 0.5 mM SUV for a PVR of 38 (330:1 lipid-peptide excess). The spectra reveal an intensity change upon binding, as spectral maxima are decreased by 25% at 5°C and by half at 75°C. Spectral red shifts are also observed with binding, with a 26nm difference in the positions of spectral maxima at 5°C. The high temperature spectrum is blue shifted by only 3 nm due both to a slight blue shift in tryptophan fluorescence upon helix unfolding and to the red shift produced by an increased population of unbound peptide at high temperature. Figure 3.4B and C show SVD bases and temperature trends, respectively, from dispersed fluorescence temperature titrations. SVD component 1 tracks the intensity change, and SVD component 2 tracks the wavelength shift. The midpoint of both intensity and spectral shift transitions are tuned to approximately 40 °C under these conditions and both are complete by 60 °C.

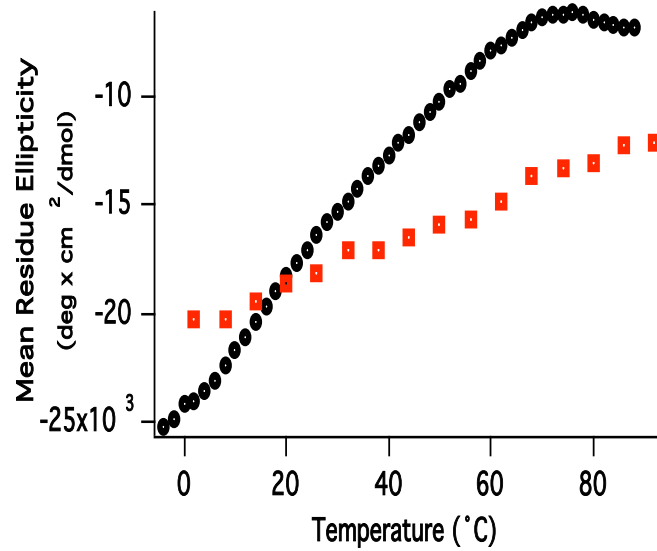
Though some of the intensity change accounted for by SVD basis and trend 1 corresponds to intrinsic tryptophan quenching at high temperature, such relaxation times are within the dead time of our temperature jump experiment so that any observed relaxation can be attributed to membrane unbinding. In addition, SVD basis 2 for Helix 5 in buffer trends in the opposite direction, so that any relaxation process producing a red shift in the data can reliably be attributed to binding events.

**Figure 3.3 A.** Mean residue ellipticity at 222nm does not change upon membrane binding at ambient temperature. **B.** Thermal titrations reveal a marked increase in helix stability upon membrane binding.

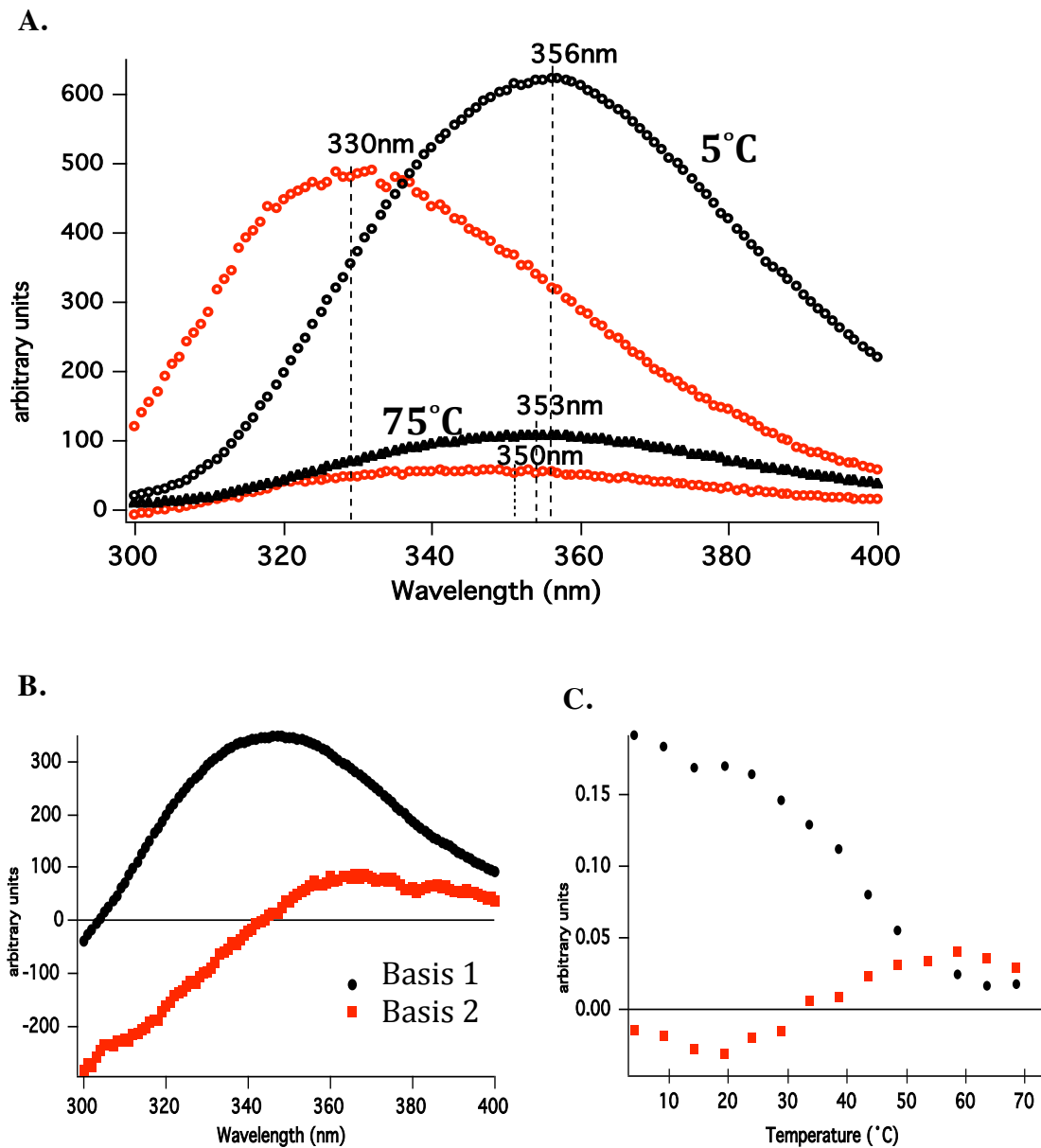
**A.**



**B.**



**Figure 3.4** Helix 5 fluorescence melt in Varian fluorometer. **A.** Background corrected spectra in the presence (red) and absence (black) of SUV. Fluorescence is quenched by 25% at 5°C and 50% at 75°C and blue shifted by 26nm at 5°C and 3nm at 75°C. **B.** SVD analysis of thermal titrations reveal two significant basis functions. **C.** The corresponding temperature dependence of those basis functions. Basis 1 (black) tracks the overall fluorescence intensity while basis 2 (red) reveals a red shift in the spectrum as the peptide unbinds from the vesicle. This cannot be due to unfolding of the peptide, since the spectral shift in the absence of vesicles trends in the opposite direction.

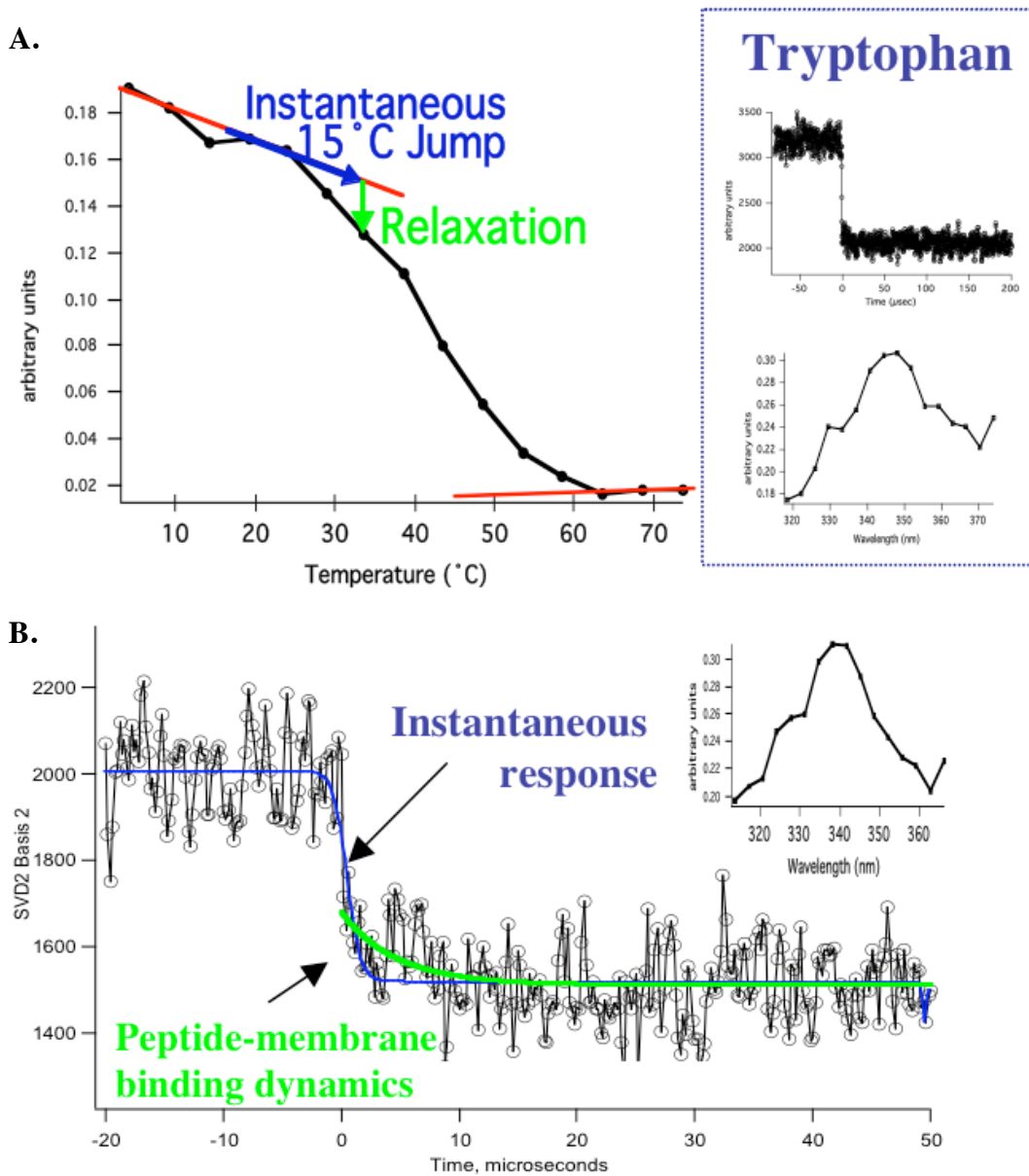


### Association/Binding Dynamics.

Figure 3.5 and 3.6 show the relaxation kinetics observed in SVD basis 1 (intensity) and basis 2 (wavelength shift) after a 16 to 34°C temperature jump on a sample of Helix 5 in SUV at a PVR of 75 and 50% D<sub>2</sub>O. Figure 3.5a and 3.6a show a schematic of the jump over the temperature trends for SVD basis 1 and 2, respectively from the equilibrium melts shown in figure 3.4. Insets show the basis functions and relaxation for Tryptophan. There is a shift in tryptophan intensity, as shown in figure 3.5a, but it is not resolved by this experiment. Figure 3.6a shows that there is no wavelength shift for tryptophan, although its SVD 2 component reveals a small, constant bias in the last PMT channel. The relaxation of SVD basis 1 for the Helix 5-SUV sample is shown in figure 3.5b. 10 jumps are averaged to achieve this signal, which is not smoothed or binned. The data is fit to a single exponential with  $\tau = 5 \pm 2$  microseconds. This timescale is consistent with the diffusion limited dissociation reaction rate,  $k_D$  expected for a 50nm vesicle and 5nm (30 residue helix) peptide.

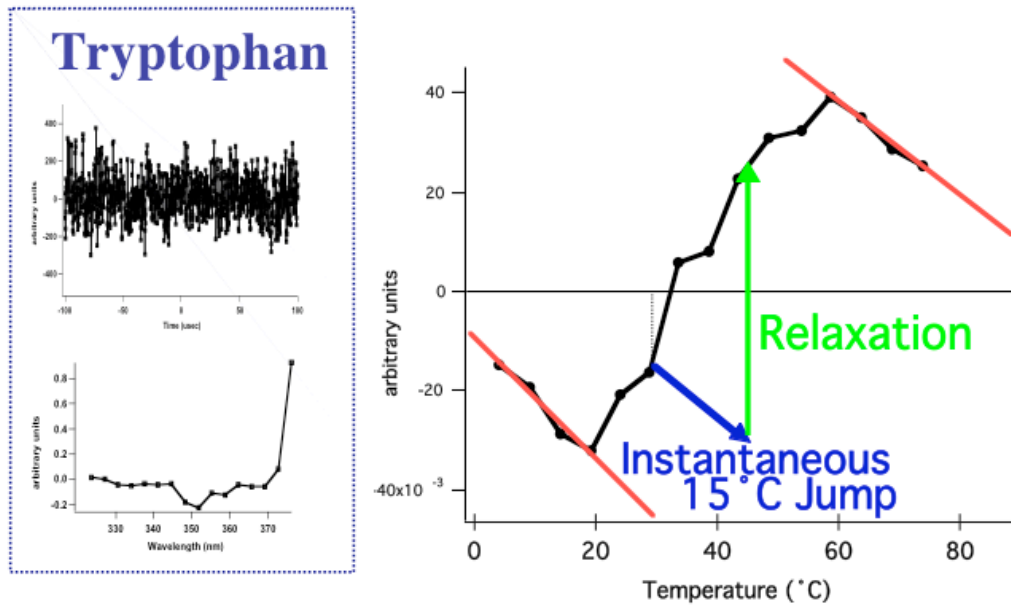
Figure 3.6b shows the relaxation kinetics observed for SVD basis 2. Since both bound and unbound baselines have negative slope as shown in figure 3.6a, a rapid initial dip in intensity is expected in this basis. This dip should be followed by a slower relaxation to the more dissociated equilibrium, and a larger SVD 2 coefficient. The kinetics trace does not show this initial dip, however it does show slower dynamics leading to a larger SVD basis 2 coefficient. A single exponential fit to this data yields a process with  $\tau = 20 \pm 5$  microseconds. This timescale is consistent with the geminate (un)embedding rate  $k_G$ .

**Figure 3.5** Relaxation kinetics monitored by SVD basis 1 (intensity). **A.** The jump and subsequent relaxation are shown schematically by blue and green arrows, respectively, drawn over the SVD basis 1 trend from figure 3.4c. The red lines in this schematic indicate the bound and unbound baseline fits. The relaxation of tryptophan alone is not resolved by this method as shown by the SVD basis and decay in the inset. **B.** The time evolution of basis 1 (inset) reveals both the instantaneous jump seen in the tryptophan sample (blue), and dynamics (green) on the order of 5 microseconds. This timescale is consistent with the diffusion limited dissociation reaction rate,  $k_D$ .

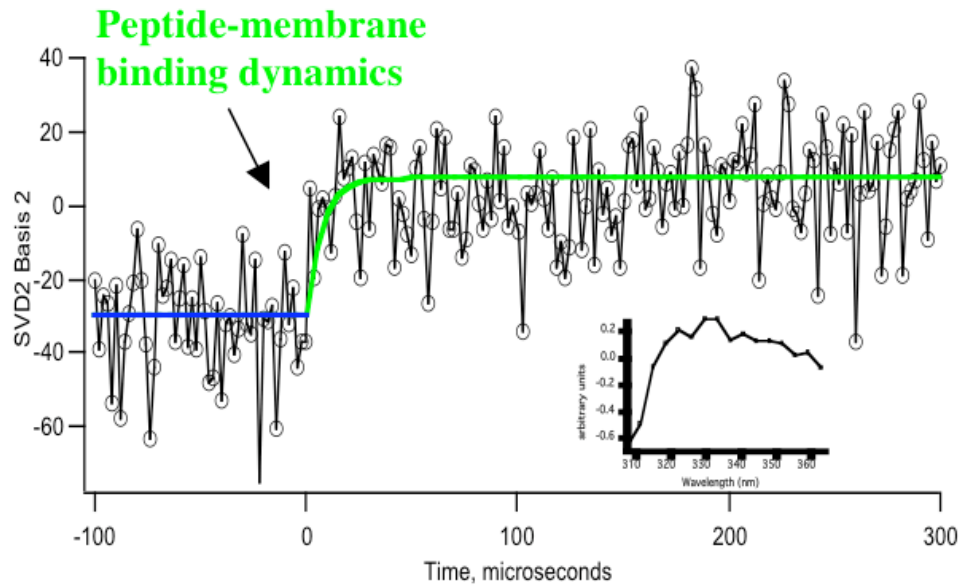


**Figure 3.6** Relaxation kinetics monitored by SVD basis 2 (wavelength shift). **A.** The jump and subsequent relaxation are shown schematically by blue and green arrows over the SVD basis 2 trend from figure 3.4c. As shown in the inset, no spectral shift is observed for tryptophan. **B.** The kinetics trace shows slower dynamics than that observed with SVD basis 1 process, on the order of 20 microseconds. This could be the geminate (un)embedding rate,  $k_G$ .

**A.**



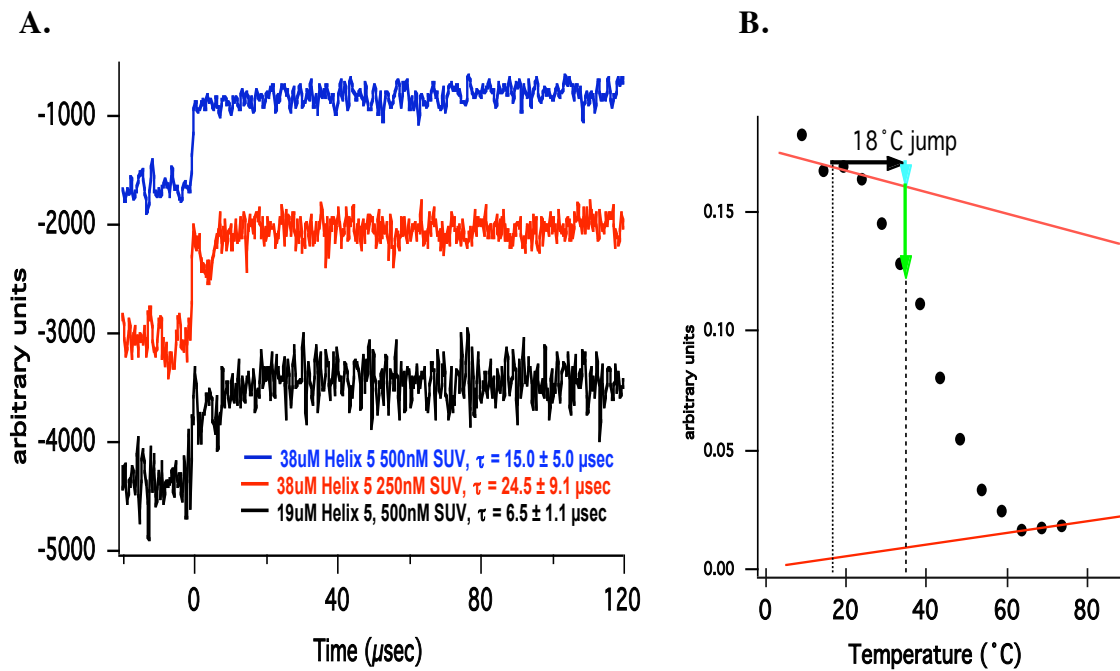
**B.**





In order to further probe whether the SVD basis 1 dynamics could correspond to the diffusion-controlled SUV-membrane association, we measured the kinetics at a range of PVR values. Specifically, we compared the rates when the peptide concentration was doubled, while keeping the vesicle concentration constant as well as the consequences of doubling the vesicle at constant protein concentration. Figure 3.7 shows the SVD basis 1 relaxation kinetics after a 16-34°C jump in samples containing a range of PVR values. These traces are averaged over 50 shots, but the data are not smoothed or binned. In agreement with first order kinetics, the rate is approximately doubled in going from 19 to 38  $\mu$ M peptide in 500nM SUV. It is unclear if there is also dependence on SUV, however, since the increased rate observed in going from 250 to 500nM SUV at 38  $\mu$ M peptide has a very large error bar.

**Figure 3.7 A.** The concentration dependence of relaxation kinetics monitored by SVD basis 1 (intensity). When the peptide concentration is doubled from 19 to 38  $\mu\text{M}$  the rate also doubles, as expected for first order kinetics. There also appears to be a rate dependence on SUV concentration, but a large amount of noise in the data makes this difficult to determine. **B.** Schematic over the SVD 1 trend from the equilibrium melt showing the 16-34°C jump (black and blue) and subsequent relaxation (green) expected when monitoring fluorescence intensity.



### 3.4 Discussion

We have proposed a 3-state mechanism by which a preformed surface helix forms a loose complex with and then binds to a membrane by embedding into the interfacial layer between the polar headgroup region and the hydrophobic membrane interior. In our measurements, we observe a fast process on the order of  $5\mu\text{sec}$  by monitoring the time course of the SVD 1 (intensity) basis function and a slower process on the order of  $20\mu\text{sec}$  by monitoring the SVD 2 (wavelength shift) basis function. These timescales are consistent with those expected for the (dis)association and geminate (un)embedding rates, respectively. In addition, the SVD 1 dynamics are consistent with the unimolecular dissociation reaction since doubling the peptide concentration at constant SUV results in a doubling of the rate. There may also be a dependence of the rate on SUV concentration, but additional measurements with improved signal to noise are required to determine this. To further determine whether this fast phase corresponds to the association step, the temperature dependence of the kinetics can be measured. For a diffusional process, this should vary as  $\eta^{-1}$ . Similarly, temperature dependent measurements can yield information about the binding step including information about the size of the barrier for this process.

These results show the first steps toward characterizing the peptide-membrane association and binding dynamics for Helix 5, but many interesting questions remain including, what is the local environment of the tryptophan? Is it buried in the hydrophobic core, as suggested by the large spectral blue shift observed with membrane

binding, or is it loosely bound in the polar headgroup region? The lack of a lifetime change, as observed for binding of Helix 5 to SUV, is more consistent with this last scenario. To shed light on this question, we are performing molecular dynamic simulations with explicit lipid and membrane in which initial conditions in which the tryptophan faces toward and away from the membrane, respectively, are being compared. If both trajectories converge to a state with an embedded tryptophan, this would give additional credence to our model and to the idea that the dynamics observed in the SVD 2 component of our time-resolved dispersed fluorescence jumps corresponds to this geminate embedding step

### 3.5 References

1. Constantinescu, I. & Lafleur, M. (2004). Influence of the lipid composition on the kinetics of concerted insertion and folding of melittin in bilayers. *Biochimica et Biophysica Acta* **1667**, 26-37.
2. Tang, J. & Gai, F. (2008). Dissecting the Membrane Binding and Insertion Kinetics of a pHLIP Peptide. *Biochemistry* **47**, 8250-8252.
3. Tang, J., Signarvic, R. S., DeGrado, W. F. & Gai, F. (2007). Role of Helix Nucleation in the Kinetics of Binding of Mastoparan X to Phospholipid Bilayers. *Biochemistry* **46**, 13856-13863.
4. Tang, J., Yin, H., Qiu, J., Tucker, M. J., DeGrado, W. F. & Gai, F. (2009). Using Two Fluorescent Probes to Dissect the Binding, Insertion, and Dimerization

- Kinetics of a Model Membrane Peptide. *Journal of the American Chemical Society* **131**, 3816-3817.
5. Tucker, M. J., Getahun, Z., Nanda, V., DeGrado, W. F. & Gai, F. (2004). A New Method for Determining the Local Environment and Orientation of Individual Side Chains of Membrane-Binding Peptides. *Journal of the American Chemical Society* **126**, 5078-5079.
  6. Russel, C. J., King, D. S., Thorgeirsson, T. E. & Shin, Y. K. (1998). De novo design of a peptide which partitions between water and phospholipid bilayers as a monomeric alpha-helix. *Protein Engineering* **11**, 539-547.
  7. Ladokhin, A. S. & White, S. H. (2004). Interfacial Folding and Membrane Insertion of a Designed Helical Peptide. *Biochemistry* **43**, 5782-5791.
  8. Reshetnyak, Y. K., Segala, M., Andreev, O. A. & Engelman, D. M. (2007). A Monomeric Membrane Peptide that Lives in Three Worlds: In Solution, Attached to, and Inserted across Lipid Bilayers. *Biophysical Journal* **93**, 2363-2372.
  9. Jayasinghe, S., Hristova, K. & White, S. H. (2001). MPtopo: A database of membrane protein topology. *Protein Science* **10**, 455-458.
  10. Gill SC, v. H. P. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **182**, 319-326.
  11. Rojas, N. L., Kamtekar, S., Simons, C. T., McLean, J. E., Vogel, K. M., Spiro, T. G., Farid, R. S. & Hecht, M. H. (1997). De novo heme proteins from designed combinatorial libraries. *Protein. Sci.* **6**, 2512-2524.

12. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109-5124.
13. Chen, Y., Mant, C. T., Farmer, S. W., Hancock, R. E. W., Vasil, M. L. & Hodges, R. S. (2005). Rational Design of  $\{\alpha\}$ -Helical Antimicrobial Peptides with Enhanced Activities and Specificity/Therapeutic Index. *Journal of Biological Chemistry* **280**, 12316-12329.
14. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *Journal of Molecular Biology* **179**, 125-142.
15. Rojas, R. (1996). *Neural Networks, A Systematic Introduction*, Springer-Verlag.
16. Park, S. H., Mrse, A. A., Nevzorov, A. A., Mesleh, M. F., Oblatt-Montal, M., Montal, M. & Opella, S. J. (2003). Three-dimensional Structure of the Channel-forming Trans-membrane Domain of Virus Protein "u" (Vpu) from HIV-1. *Journal of Molecular Biology* **333**, 409-424.
17. Wimley, W. C. & White, S. H. (2000). Designing Transmembrane R-Helices That Insert Spontaneously. *Biochemistry* **39**, 4432-4442.
18. Liu, L.-P. & Deber, C. M. (1997). Anionic Phospholipids Modulate Peptide Insertion into Membranes. **36**, 5476-5482.
19. Chang, G. (2003). Structure of MsbA from *Vibrio cholera*: A Multidrug Resistance ABC Transporter Homolog in a Closed Conformation. *Journal of Molecular Biology* **330**, 419-430.

20. Deisenhofer, J., Epp, O., Sinning, I., Michel, H. (1995). Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodospseudomonas viridis*. *Journal of Molecular Biology* **246**, 429-457.
21. Harrenga, A. & Michel, H. (1999). The Cytochrome C Oxidase from *Paracoccus Denitrificans* Does not Change the Metal Center Ligation Upon Reduction. *Journal of Biological Chemistry* **274**, 33296.
22. Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E. M., Pebay-Peyroula, E. & Navarro, J. (2001). X-Ray Structure of Sensory Rhodopsin II at 2.1 Å Resolution. *Proceedings of the National Academy of Science (PNAS)* **98**, 10131-10136.
23. Tisa, L. S. & Rosen, B. P. (1990). Molecular characterization of an anion pump. The ArsB protein is the membrane anchor for the ArsA protein. *Journal of Biological Chemistry* **265**, 190-194.
24. van den Berg, B., Clemons Jr., W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44.
25. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. (2003). Structure and Mechanism of the Glycerol-3-Phosphate Transporter from *Escherichia Coli*. *Science* **301**, 616-620.
26. Waters S.H., R. P., Grinsted J., Altenbuchner J., Schmitt R. (1983). The tetracycline resistance determinants of RP1 and Tn1721: nucleotide sequence analysis. *Nucleic Acids Research* **11**, 6089-6105.

## Section II

### Effects of Macromolecular Crowding on Protein Folding and Stability

#### Introduction

Though most in vitro measurements of biological reactions are performed in dilute solution, the cell cytoplasm and interstitial space contains 50-400g/L of protein, while cell nuclei contain up to 1m of DNA in addition to RNA and DNA binding proteins. At these concentrations, it is estimated that macromolecules account for up to 40% of the cell's volume.<sup>1;2</sup> This leads to a wide range of non-specific interactions, such as electrostatic repulsion and attraction, hydrophobic attraction, and steric repulsion, also known as the excluded volume effect that contribute to the overall energetics. These interactions can significantly alter reaction rates and equilibria in the crowded cellular environment, though they are not well understood.

Consider, for example, how excluded volume affects the chemical potential.<sup>3</sup> When calculating energies and equilibrium constants in dilute solution, the assumption of ideality is usually valid and concentrations can be used directly. However, under crowded conditions, the available volume for molecular species  $i$  is significantly reduced, resulting in an activity coefficient greater than 1,

$$\gamma_i = \frac{a_i}{c_i} = \frac{v_{total}}{v_i} > 1 \quad (1)$$

where  $\gamma_i$  and  $c_i$  are the activity coefficient and concentration, respectively, and  $a_i$  is the thermodynamic activity which contributes to the overall energy:



$$\mu = \mu_i^0 + RT \ln(a_i) = \mu_i^0 + RT \left( \ln c_i + \ln \left( \frac{v_{total}}{v_i} \right) \right) \quad (2)$$

where the first two terms give the energy in dilute solution and any decrease in available volume,  $v_i/v_{total}$ , for species  $i$  corresponds to an increase in energy.

If in dilute solution we can approximate  $v_i \approx v_{total}$ , then the energy change associated with a given excluded volume fraction,  $\phi_c$ , is,

$$\Delta\mu^{\phi-0} = RT \cdot \ln(1 - \phi_c) = RT \ln(1 - c_c \cdot v_c) \quad (3)$$

where  $v_c$  and  $c_c$  are the molecular volume and concentration of the crowding agent. It is clear from (3) that the energy increases with increasing excluded volume.

For protein folding, the volume contraction in going from the unfolded to the folded state means that  $(v_F/v_U) > 1$ . This means that an increase in stability is expected to occur in the presence of crowding agents according to:

$$\Delta\Delta G_{F-U}^{\phi-0} = \Delta\mu_F^{\phi-0} - \Delta\mu_U^{\phi-0} = RT \ln \left( \frac{v_U^{\phi}}{v_F^{\phi}} \right) \quad (4)$$

Similarly, the transition state is much more compact than the unfolded state and so  $(v_{TS}/v_U) > 1$ , where  $v_{TS}$  refers to the available volume in the transition state. The change in activation energy is given by:

$$\Delta\Delta G_{TS-U}^{\phi-0} = \Delta\mu_{TS}^{\phi-0} - \Delta\mu_U^{\phi-0} = RT \ln \left( \frac{v_U^{\phi}}{v_{TS}^{\phi}} \right) \quad (5)$$

which is reduced by the presence of crowding agents and so an increase in the forward folding rate is expected. By similar logic,  $\Delta\Delta G_{F-TS}^{\phi-0}$  is expected to be larger and the unfolding rate slower, since  $(v_{TS}/v_F) < 1$ ; however, this effect is not expected to be as large

since there is typically a much larger compaction in going from the unfolded to transition state, than from the transition state to the folded conformation.

In general,  $v_1 \neq v_{\text{total}}$  in dilute solution and the available volume for a given crowding fraction is difficult to calculate, particularly for the unfolded state, since it is not a single, well-defined state, but an ensemble of states with distinct conformations and the number of molecules in each state is hard to quantify.

An easier way to estimate the energy contribution of the excluded volume effect is to use conformational entropy. The difference between folded and unfolded state's crowding-induced change in conformational entropy is related to the energy change by:

$$\Delta\Delta G_{F-U}^{\phi-0} = \Delta\Delta H_{F-U}^{\phi-0} - T\Delta\Delta S_{F-U}^{\phi-0} = T(\Delta\Omega_U^{\phi-0} - \Delta\Omega_F^{\phi-0}) \quad (6)$$

There is no change in the folding enthalpy,  $\Delta H_{F-U}^{\phi-0}$ , since the excluded volume effect is purely steric and does not include specific interactions. It is clear from (6) that the crowding induced change in protein stability depends only on the difference in the reduction of the number of possible conformations for the folded state,  $\Delta\Omega_F^{\phi-0}$ , and the unfolded state,  $\Delta\Omega_U^{\phi-0}$ , in the presence and absence of crowding agents. Since the folded protein is much more compact than the unfolded state, the reduction in the number of conformations associated with an excluded volume fraction  $\phi_c$  will be less. In other words,  $\Delta\Omega_F^{\phi-0} > \Delta\Omega_U^{\phi-0}$  where  $\Delta\Omega_F^{\phi-0} \leq 0$  and  $\Delta\Omega_U^{\phi-0} < 0$ .

Using an off-lattice model in a continuum solvent and hard sphere crowding agents of the same radius as the (folded) test protein, Cheung et al<sup>4</sup> calculated the folding free energy and activation energy at a range of crowding fractions, from 0 to 0.25.

According to their calculations, the percent change in the melting temperature,  $T_m$ , is related to the crowding fraction by  $\Delta T_F = 0.84\phi_c^{1.8}$ . At the highest crowding fraction used in their calculations,  $\phi_c=0.25$ , they saw an approximately 1kCal/mol increase in stability for the protein studied (WW domain). They also saw a non-monotonic dependence of the rate with crowding fraction, wherein the maximum, 3-fold increase, is reached at a crowding fraction of 10%. These calculations offer quantitative predictions that can be compared with folding experiments in the presence of crowding agents. Though some studies have demonstrated agreement with these predictions,<sup>5; 6; 7; 9</sup> there are still questions that remain. Most significantly, since most studies are performed using branched polymers, like PEG, Dextran and Ficoll, it is unclear whether their results are valid given the requirement that the excluded volume effect depends on the mutual impenetrability of crowding agents with one another and with the protein under study in the manner of a hard sphere. Ficoll 70 is the most promising of the branched polymers, with a Stokes radius of 5.1nm and a molecular weight of 70k/mol, but this still be too porous. An average cellular protein has a Stokes radius of 2.24nm and a density of 1.42 g/ml<sup>6</sup> and densities are not much less than this for even the most loosely packed proteins. So, with a 0.2g/ml molecular density, Ficoll density is 70 times less than that of proteins, which are the dominant *in vivo* crowding agents. Some experimental data already suggests that interactions other than excluded volume may be required to explain the effect of high Ficoll concentrations on proteins. Zhou et al recently showed<sup>7</sup> that when BSA is combined with Ficoll 70 or Dextran 70 at a total concentration of 100g/L and a 1:9 ratio of protein to sugar, the kinetic effects on lysozyme refolding are non-additive. The yields were greater by about 20% and 10% compared to protein-only crowding or sugar-only

crowding measurements, respectively. Further, Ficoll 70 yields were about twice that of BSA only crowding, indicating that these two crowding agents have distinct effects on lysozyme stability.

In this section, I attempt to characterize the effects of excluded volume on the stability and kinetics of a stable Lambda Repressor mutant Y22WQ33YA3749G. I first characterize the folding reaction in buffer and then compare this to that in solutions of >10% w/v Ficoll 70. I also compare Ficoll measurements to those in the presence of sucrose to determine whether non-specific interactions other than steric repulsion are involved. I then develop a protein crowding system amenable to thermally-induced kinetic and stability measurements of folding, which can be directly compared to measurements in Ficoll.

## References

1. Ellis, R. J. M., A. P. (2003). Plugging the transport gap. *Nature* **425**, 27-28.
2. Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Science* **26**, 597-604.
3. Chebotareva, N. A., Kurganov, B. I., and Livanova, N. B. (2004). Biochemical Effects of Molecular Crowding. *Biochemistry (Moscow)* **69**, 1239-1251.
4. Cheung, M. S., Klimov, D. & Thirumalai, D. (2005). Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins. *Proceedings of The National Academy of Sciences* **102**, 4753-4758.

5. Perham, M., Stagg, L. & Wittung-Stafshede, P. (2007). Macromolecular Crowding Increases Structural Content of Folded Proteins. *FEBS Letters* **581**, 5065-69.
6. Fischer, H., Polikarpov, I. & Craievich, A. F. (2004). Average protein density is a molecular-weight-dependent function. *Protein Science* **13**, 2825-2828.
7. Zhou, B., Yi, L., Fen, D., Zheng, Z. & Jie, C. (2004). Mixed macromolecular crowding accelerates the oxidative refolding of reduced, denatured lysozyme: implications for protein folding in intracellular environments. *Journal of Biological Chemistry* **279**, 55109-55116.

## Chapter 4

### Crowding Studies of a Downhill Folder

#### 4.1 Introduction

This study presents the characterization of the thermodynamics and kinetics of Lambda Repressor mutant Y22WQ33YA3749G in the presence and absence of crowding agents. These mutants are used to address two important questions: 1). Are the effects of Ficoll different from the effects of small osmolytes such as sucrose? Sucrose is a known cosmotrope which increases protein stability by altering water structure<sup>1; 2; 3; 4; 5</sup>. This is contrasted with the steric repulsion responsible for the excluded volume effect, which is expected to be minimal for sucrose due to its extremely small size compared to proteins<sup>2; 3; 4; 5</sup>. If sucrose solutions also produce shifts in the stability and folding rates of proteins, can these effects be distinguished from the excluded volume effect purported to account for shifts observed with the same volume fraction of Ficoll? 2). What are the effects of crowding on downhill protein folders? Crowding theory predicts that, while forward reaction rates are increased for transition-state limited protein folders, forward rates for downhill folders are predicted to decrease, while stability is increased in both cases<sup>3</sup>. This is because destabilizing the unfolded state cannot result in a lower barrier if no barrier exists, but it will still contribute to increasing the energy difference between unfolded and folded states. A decrease in the folding rate is further expected when the barrier is lowered enough so that diffusion-limited folding becomes dominant. This is

because of the high viscosities of crowded solutions, rather than steric repulsion characteristic of macromolecular crowding.

## 4.2 Materials and Methods

### Sample preparation.

Lambda repressor mutant Y22WQ33YA3749G, henceforth referred to as  $\lambda$ nQ33Y, is expressed in a PET-19b vector (Novagen, San Diego, CA) and expressed in Rosetta (DE3) pLysS (Novagen) cells in media containing 20 g/L tryptone, 10 g/L yeast-extract, 5 g/L NaCl, 200 mg/L ampicillin, and 4 g/L glucose at pH 7.4. Cell culture was scaled up from 20mL to 12L and induced using 2 mM isopropyl- $\beta$ -Dthiogalactopyranoside once the cell density was sufficient to produce an OD of 0.8-1 at 600nm. Induced cultures were kept in a shaker overnight at room temperature ( $\sim 20^{\circ}\text{C}$ ) before harvesting via centrifugation and lysis by passing the sample twice through a French press at 12,000 psi. Purification was performed using a Ni-NTA column (QIAGEN, Tokyo, Japan) with imidazole as the eluting reagent. Histidine tags were removed by incubating samples for  $\sim 24$  hours at  $20^{\circ}\text{C}$  with  $\sim 1$  Unit of thrombin (Novagen) per mg of protein and the final purification was achieved using a Sephacryl S-200 HR (Amersham Biosciences, Uppsala, Sweden) size-exclusion column using 50 mM  $\text{NaH}_2\text{PO}_4$ /500 mM NaCl run buffer at pH 7. Purified protein is then dialyzed against double deionized water, lyophilized and stored at  $70^{\circ}\text{C}$  prior to use.

All experiments are performed in pH 7 phosphate buffer. The desired volume fraction of Ficoll 70/sucrose is achieved by first adding half of the total volume of buffer

and then adding Ficoll 70 or sucrose until the volume is raised by the desired amount and then again adding buffer to the desired final volume. The known molecular density of 0.2g/ml, 1.4g/ml and 0.6g/ml is used to estimate the mass of Ficoll, protein, and sucrose, respectively, required to achieve the desired volume fraction.

### **Thermodynamic measurements.**

Steady state circular dichroism and fluorescence intensity measurements were carried out in a Jasco (Easton, MD) J-715 equipped with a Peltier temperature controller (Jasco) and converted to mean residue ellipticity as described previously in chapter 3. Protein thermal denaturation curves are fit to a two-state model with linear folded and unfolded baselines. CD spectra, integrated fluorescence and time-averaged 222nm CD signal are all collected as samples are heated from 10 to 90°C. All samples were checked for hysteresis by repeating measurements after overnight storage at 4°C.

### **Determination of folding rates.**

Temperature jump folding kinetics<sup>6</sup> were induced by a 10 ns Raman-shifted Nd:YAG laser pulse as described in chapter 3. Folding was probed by a continuous pulse train of 280 nm, 200 fs duration. Ti:sapphire laser pulses spaced by 14 ns to excite tryptophan 22, tyrosine 33, and tyrosine 60. Changes in the fluorescence emission lifetime were used to track the protein folding kinetics. These transients were averaged using a bin size beginning at 280nanoseconds and increasing logarithmically with time. These decays were fit to double-exponential decays in which the first time constant is fixed at 1 μsecond in accordance with previously measured values<sup>7</sup> for the “molecular



phase” time constant,  $\tau_m$ .

### Viscosity measurements.

In order to accurately compare folding rates in crowded solution and buffer, I measured the viscosity of a 22% w/v Ficoll solution in 50mM phosphate buffer. I performed these measurements using a Carri-Med CSL<sup>2</sup> 500, controlled stress rheometer (TA Instruments, New Castle, DE) using a 2 cm parallel plate geometry with a solvent trap and a peltier heater. Measurements were performed at 20, 45, 48, 50, 53, 55, 58, 60 and 63°C. I scanned at a 70MHz shear rate and average for 145 seconds, using a water-filled frictionless trap to prevent solvent evaporation and a thermocouple placed directly into the sample to monitor the temperature prior to each measurement. These measurements are scaled according to the known value at 20°C (Amersham Biosciences data file) and then globally fit together with water viscosities from the literature<sup>8</sup>. Similar fits were performed for sucrose together with water measurements from the literature<sup>8</sup> and the best fits to viscosity as a function of concentration, X, in w/v units and temperature, T, in degrees Celcius, are:

Sucrose

$$\eta(T, X) = (1 + 0.669X - 0.0486X^2 + 0.00089X^3) \cdot (1 + (1.474 + 0.012X) \cdot e^{-0.045T}) + (-0.653 - 0.0059X) \cdot e^{-0.00045T} \quad (1)$$

Ficoll

$$\eta(T, X) = (1 - 0.3843X + 0.00508X^2) \cdot (1 + (1.966 + 0.234X) \cdot e^{-0.0235T}) + (-1.275 + 0.0283X) \cdot e^{-0.0036T} \quad (2)$$

Water

$$\eta(T, X) = (0.226 + 1.0723 \cdot e^{-((T-10)/33)}) \quad (3)$$

These expressions are used to scale observed rates in order to distinguish viscosity and excluded volume effects.

## 4.3 Results

### Thermodynamic Measurements.

Figure 4.1 shows the CD spectra at 20°C for  $\lambda$ nQ33Y in buffer sucrose and Ficoll. The protein is highly helical in solution, however addition of 16% w/v Ficoll or sucrose reduces the signal at 222nm by more than half. This reduction in signal is less pronounced at high temperature, as shown in figure 4.2a in which CD signal at 222nm is plotted as a function of concentration at 20°C, 40°C, 60°C and 80°C. The concentration dependence of 222nm CD signal is fairly flat for all non-zero concentrations measured, with the most dramatic jump occurring between 0 and 6%, the lowest concentration measured. Figure 4b shows the spectra from 205 to 245nm revealing alpha helical signatures at low temperature in all samples. It is interesting to note the differences in the  $CD_{222}/CD_{206}$  ratio for the three samples. This ratio is often used to determine the extent of folding, the larger value indicating a greater value for the folding equilibrium constant. The largest 222nm signal is achieved in buffer, but the largest ratio is given by the sucrose sample, the only one which produces a ratio larger than one at 20°C. The ratios

for  $\lambda$ nQ33Y in sucrose, Ficoll and buffer at 20°C are 1.02, 0.88 and 0.94 respectively.

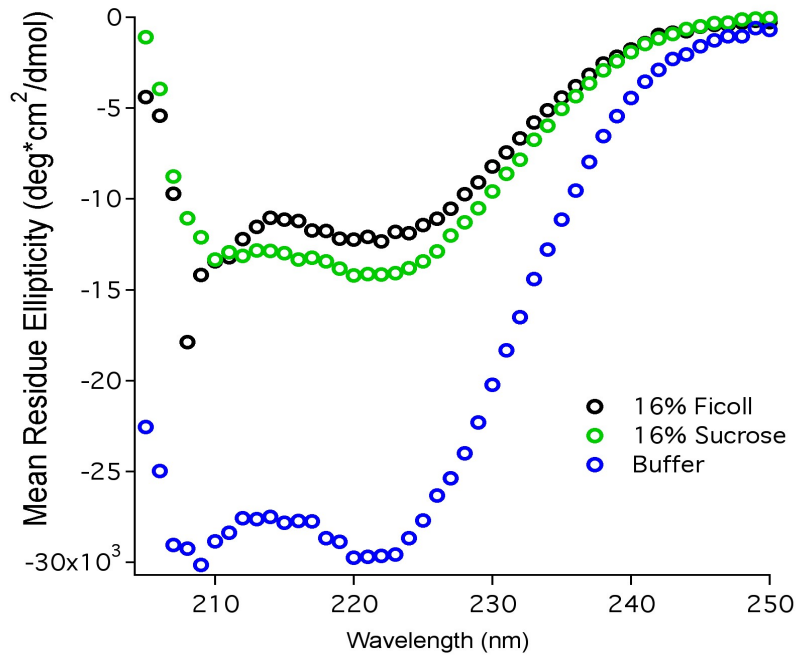
The sucrose sample continues to exhibit a  $CD_{222}/CD_{206}$  ratio greater than 1 for all temperatures measured (up to 80°C). This indicates that  $\lambda$ nQ33Y in sucrose is stabilized relative to buffer, while the protein is destabilized in Ficoll. This is not surprising as cosmotropes like sucrose are known to protect proteins against thermal denaturation. Figures 4.3 and 4.4 show the CD and fluorescence thermal titration results, respectively.

These melts give a somewhat different picture of the effects of sucrose and Ficoll on  $\lambda$ nQ33Y stability. The top panel of figure 4.3 shows the measured CD melts in buffer and 6%, 10.5%, 13%, 16%, and 22% w/v sucrose or Ficoll along with the fits used to calculate the native fraction shown in the bottom panels. No discernable difference exists among all the concentrations or between sucrose and Ficoll, and the calculated  $\lambda$ nQ33Y  $T_m$  values are all between 53-55°C. The only exception is the 10.5% Ficoll sample, shown in the inset, in which the  $T_m$  is reduced from 54°C in buffer or sucrose to 51.5°C in Ficoll.

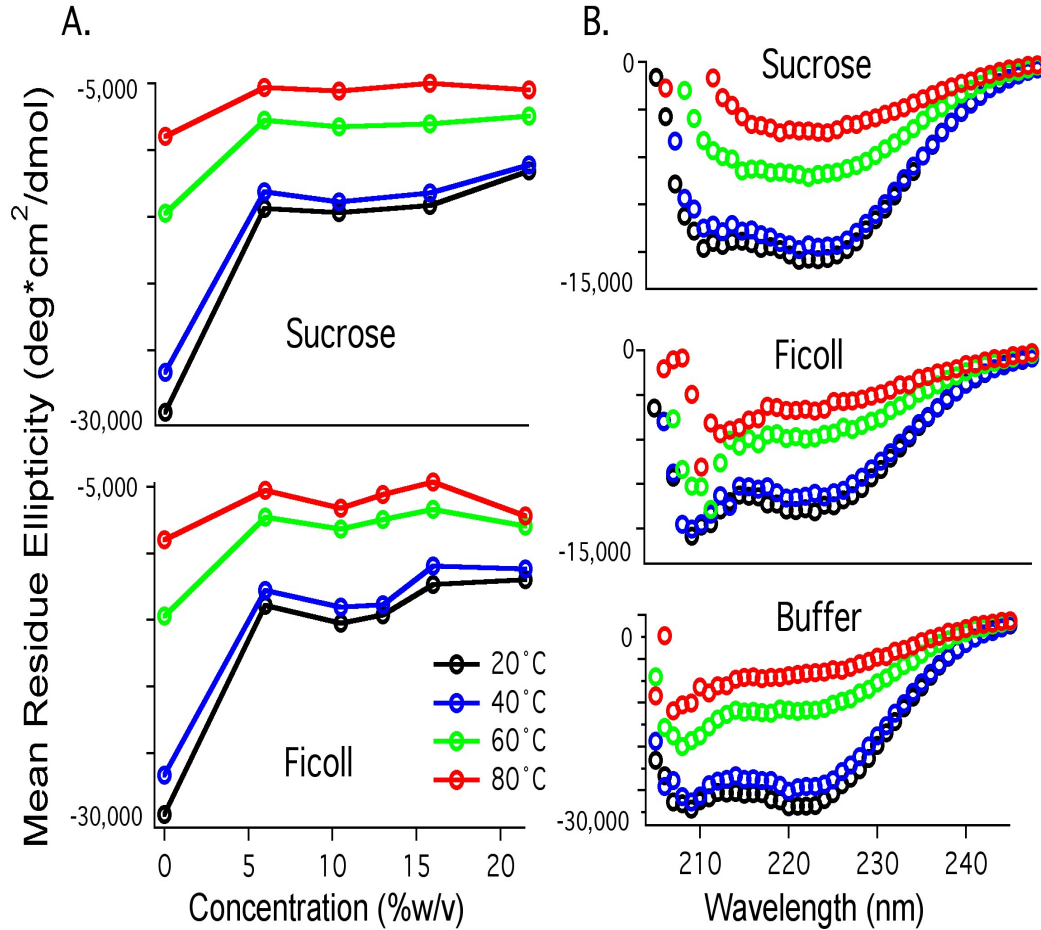
Fluorescence melts shown in figure 4.4 present a very different picture of stability from the CD measurements. Melts are shown for protein in 0, 10, 15 and 20% Ficoll and 15% w/v sucrose. All samples containing Ficoll have indistinguishable curves and  $T_m=54^\circ\text{C}$  in agreement with values measured by CD. Protein in buffer, however, appears destabilized relative to these samples, revealing a  $T_m=49^\circ\text{C}$ . Protein in 15% sucrose is also destabilized, with a  $T_m=51^\circ\text{C}$ , and is also much less cooperative, as revealed by the broader transition. This loss of cooperativity is in agreement with CD spectra from figure 4.2, which showed a reduced  $CD_{222}$  compared to protein in buffer at low temperature, but

continued to exhibit a helical CD signature at 80°C when buffer and Ficoll samples were denatured.

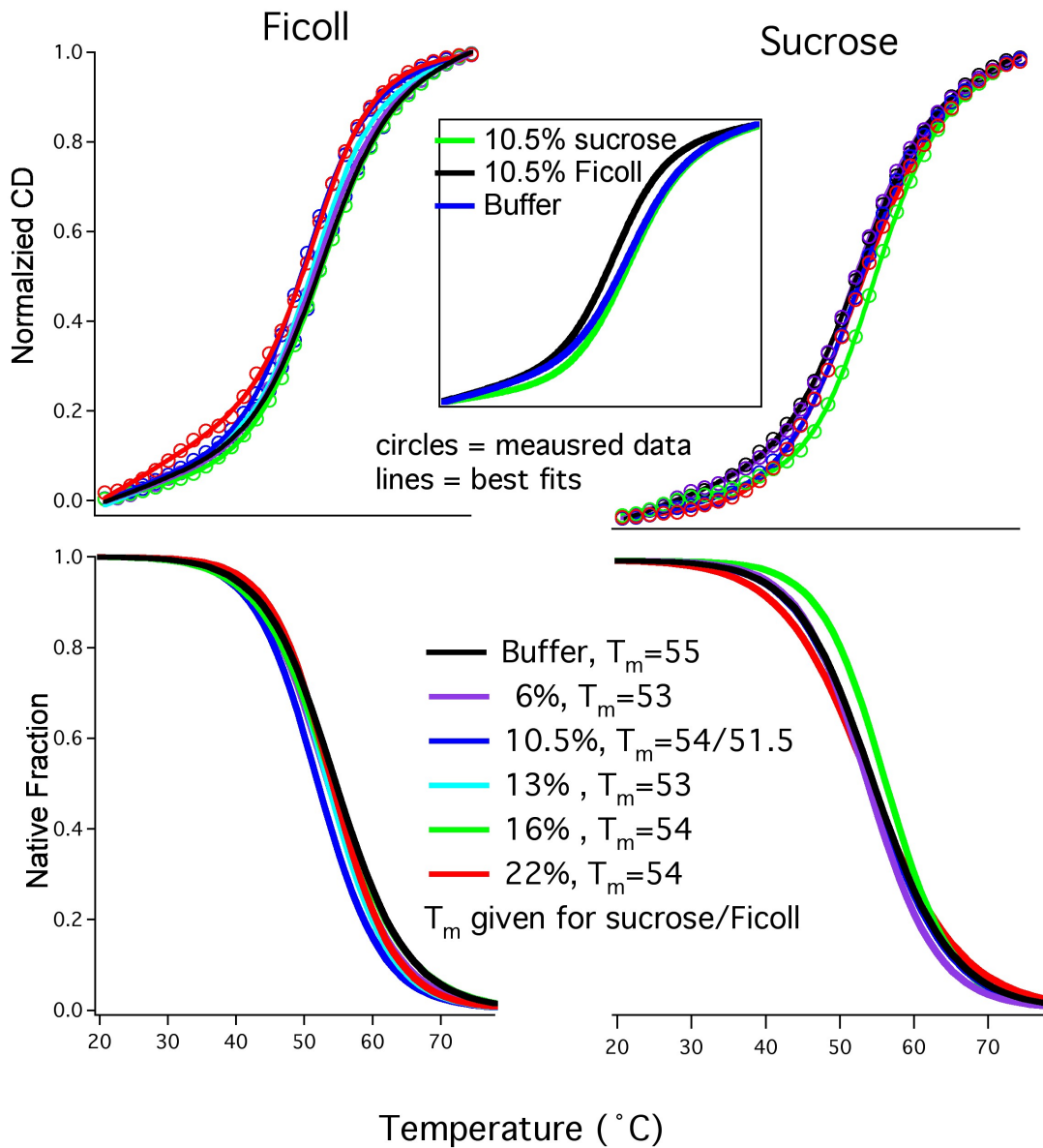
**Figure 4.1** CD spectra for Y22WQ33YA3749G in the presence of 16% Ficoll (black), 16% sucrose (green) and phosphate buffer only (blue). Addition of crowding agents significantly reduces the CD signal indicating a loss of secondary structure.



**Figure 4.2 A.** The reduction in  $CD_{222}$  is seen even at 6% crowding agent and is fairly flat over the concentration 6-22% w/v. This is true at temperatures of 20 - 80°C. **B.** CD spectra at 20°C are characteristic of the highly helical lambda repressor, but loss of structure is clear, particularly in the Ficoll sample which has the lowest  $CD_{222}/CD_{206}$  ratio. Sucrose spectra reveal significant structure is retained even at 80°C in keeping with its role as a cosmotropic protein stabilizer.

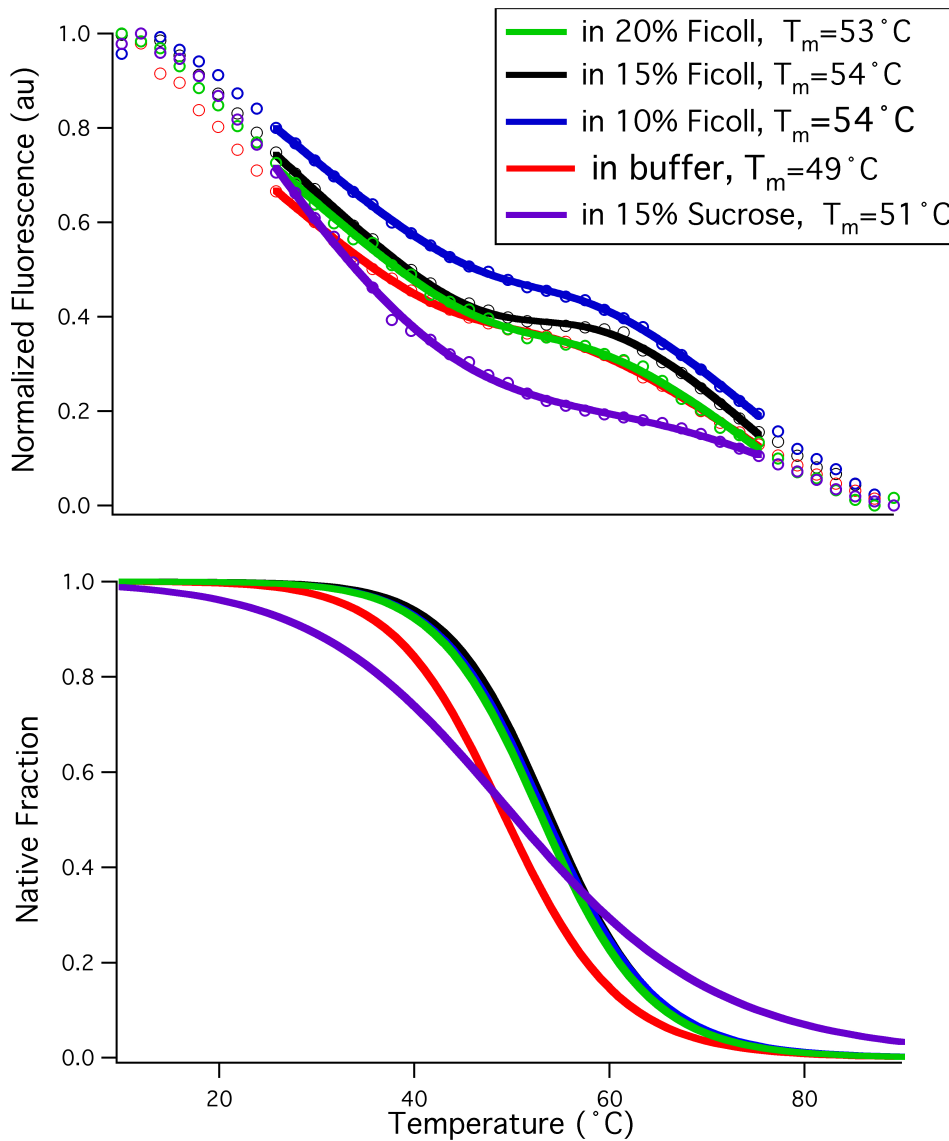


**Figure 4.3** CD melts in 6-22% Ficoll and sucrose are compared with that in buffer. The top panels show the raw melt data (open circles) together with the measure fits (lines) that were used to calculate the native fraction plots and the  $T_m$  values shown in the bottom panel. There is no discernable difference in shape or  $T_m$  for these samples, with the exception of the 10.5% Ficoll sample (top panel, inset) in which the  $T_m$  is lowered from 54°C to 51°C.



**Figure 4.4** Fluorescence melts for 0, 10, 15 and 20% Ficoll as well as 15% sucrose.

Melt data (open circles) and fits (lines) are shown in the top panel and the native fraction plots used to compute  $T_m$  are shown in the bottom panel. The protein in buffer is destabilized relative to all non-zero sugar solutions, with a  $T_m$  of 49°C. The protein in sucrose is also slightly destabilized, with  $T_m$ =51°C and (un)folding is less cooperative than any of the other samples.

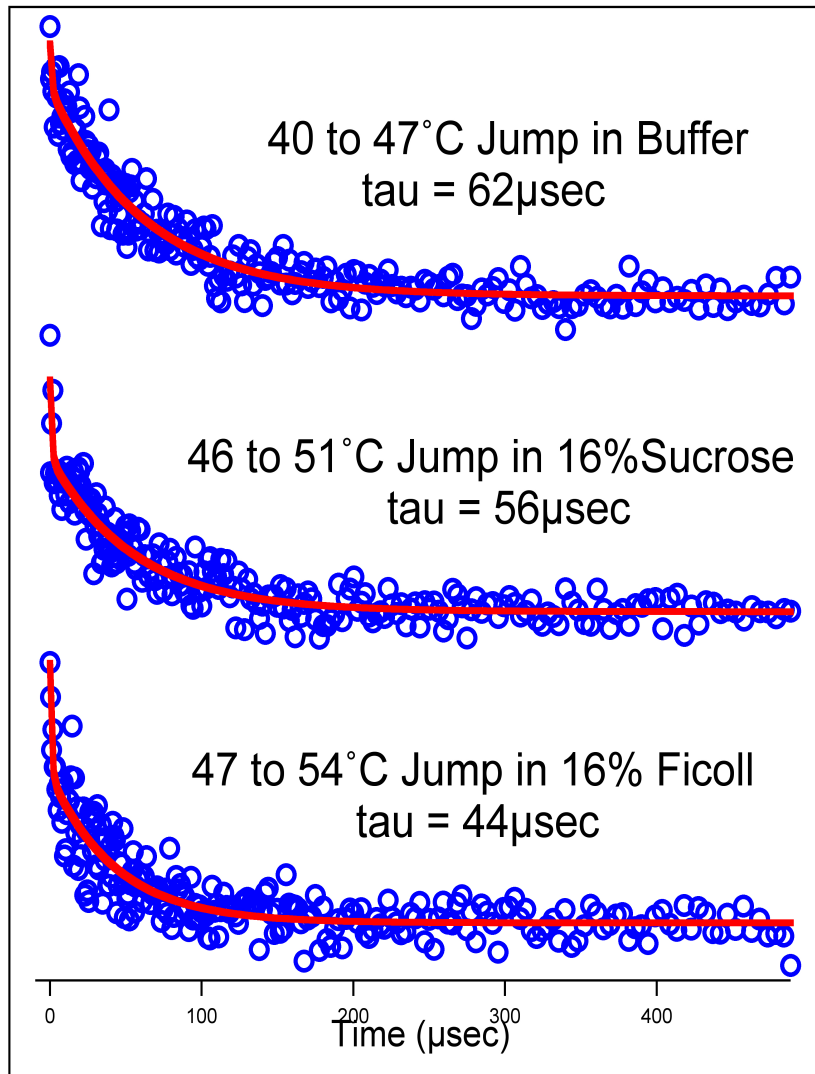


## Kinetics Measurements.

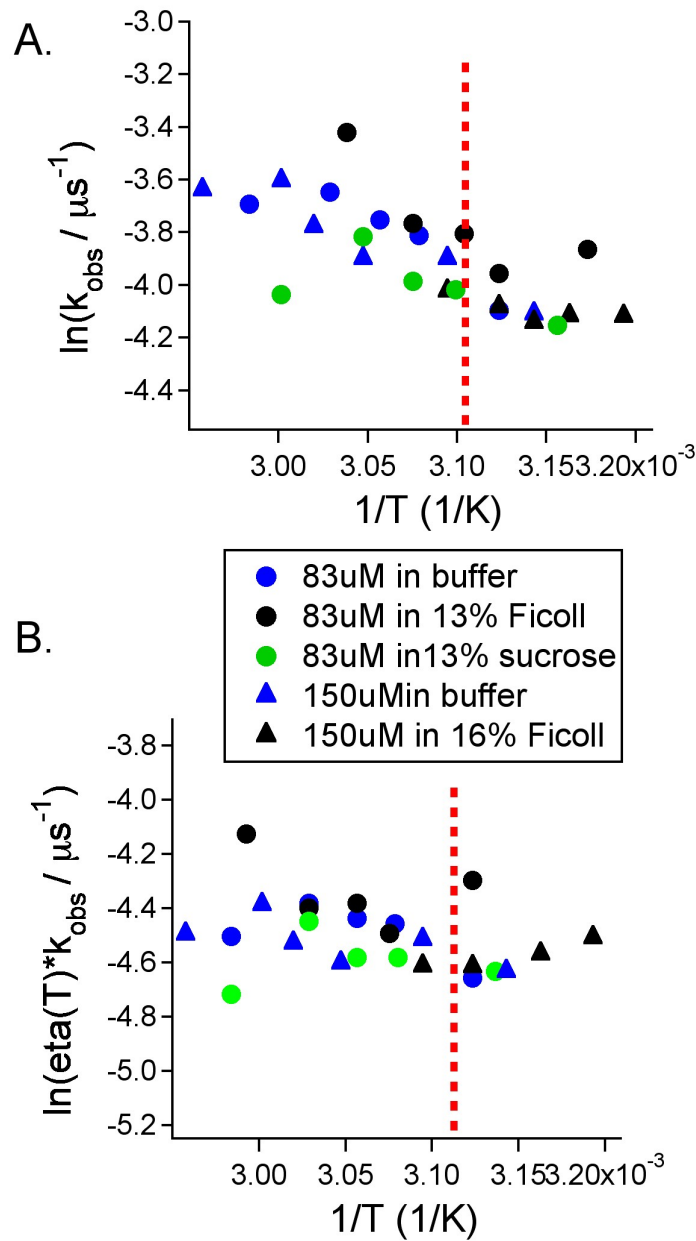
Figures 4.5 – 4.7 show the relaxation kinetics for  $\lambda$ nQ33Y near the transition temperatures shown in figure 4.4. Both 83 $\mu$ M and 150 $\mu$ M samples were measured. The lower concentration was measured in buffer alone and in 10.5% and 16% w/v Ficoll and sucrose. The higher concentration was measured at 16% Ficoll and in buffer. Figure 4.5 shows the logarithmically binned kinetic traces near the  $T_m$  (measured by fluorescence) for the 83 $\mu$ M samples in buffer, 16% Ficoll and 16% sucrose. The red curve is the double exponential fit that was used to determine activated rates. The first “molecular” phase rate,  $k_m$ , is fixed at 1 $\mu$ second in all cases as described previously<sup>7</sup>. At these temperatures,  $\lambda$ nQ33Y appears to be fastest in Ficoll and slowest in buffer. Since no differences were observed among the 13% and 16% concentrations for Ficoll or sucrose, the data were averaged to provide smoother data for the Arrhenius plots shown in figures 4.6 and 4.7. Figure 4.6 shows the  $k_{obs}$  both with (fig 4.6B) and without (fig 4.6A) viscosity scaling. That this scaling removes the slight temperature dependence observed, indicates that diffusion controlled processes play a dominant role. The folding rate Arrhenius plots shown in figure 4.7 have roughly the same positive linear temperature dependence with (fig 4.7B) and without (fig 4.7A) viscosity scaling. No difference is observed in folding rates for  $\lambda$ nQ33Y in buffer and in 13% Ficoll.



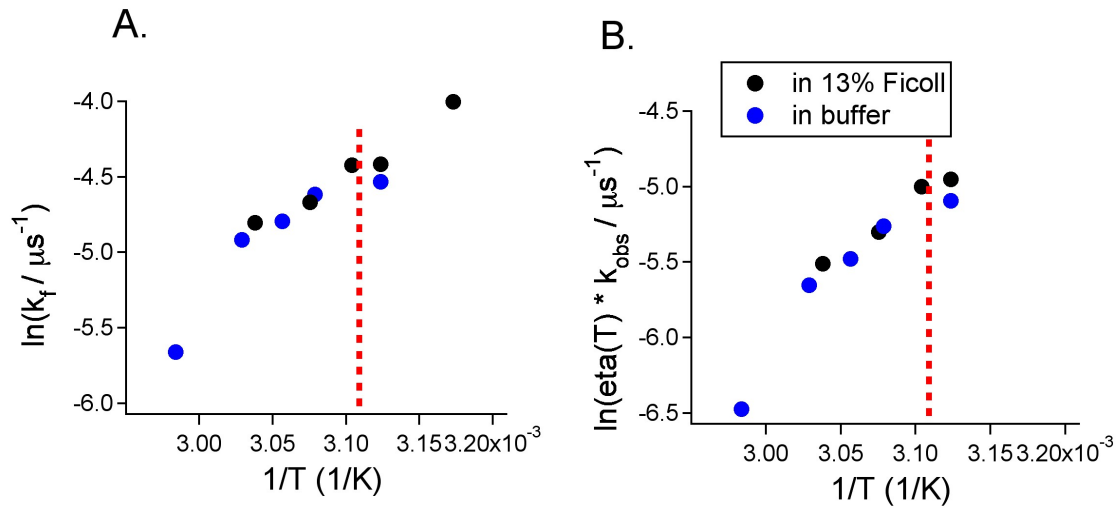
**Figure 4.5** Kinetic traces for  $\lambda$ nQ33Y in buffer (top), 16% sucrose (middle) and 16% Ficoll (bottom) near their folding midpoints as measured by fluorescence intensity. Despite having the largest viscosity, the unscaled  $k_{\text{obs}}$  for  $\lambda$ nQ33Y in Ficoll at this temperature is faster than in sucrose or buffer.



**Figure 4.6** Arrhenius plots for folding rates measured at either 83 $\mu$ M or 150 $\mu$ M  $\lambda$ nQ33Y in 13% or 16% Ficoll (black), 13% sucrose (green), and buffer (blue). All curves are shifted so that their  $T_m$  values coincide with that of the  $\lambda$ nQ33Y in buffer (49°C), indicated by the red line. A. Unscaled observed rates do not differ for the various conditions, but a slight temperature dependence is observed. B. Scaling the rates by their viscosity removes the temperature dependence, indicating that diffusion processes are dominant.



**Figure 4.7** Arrhenius plots of the folding rate for 83 $\mu$ M  $\lambda$ nQ33Y in buffer and 13% Ficoll. Rates with and without crowding agents do not differ for these two samples either before (A) or after (B) viscosity scaling. Rates exhibit the same linearly increasing temperature dependence.



## 4.4 Discussion

When performing measurements in crowded solutions, viscogenic effects can have large effects on folding rates. It has been shown previously for similar mutants of lambda repressor that, while the activated rate is insensitive to viscosity, the fast “molecular” rate scales inversely with viscosity<sup>6; 8; 9</sup>. At the high ( $\sim 20\%$  w/v) concentrations of Ficoll and sucrose used in most crowding studies, the viscosity is 10 and 2 times that of water, respectively, at 20°C. In order to distinguish steric crowding from viscosity effects, and to accurately compare crowding effects on energy barriers, it

is important to scale observed measurements by the viscosity at the crowding fraction and temperature used.

The viscosity dependence of protein folding rates stems from the diffusive nature of the associated barrier-crossing events, namely, from the requirement of polypeptide chain motion to form the stabilizing native contacts via diffusion. When these diffusive motions are dominant, the temperature dependence of reaction rates can be accounted for entirely by the temperature dependence of viscosity. This was observed for  $\lambda$ nQ33Y, which is a very fast folder which likely folds downhill. Further evidence that  $\lambda$ nQ33Y is a downhill folder are the steady-state fluorescence and CD temperature titrations. In buffer, the  $T_m$  of  $\lambda$ nQ33Y measured by these two methods differs by 5°C. This probe-dependent behavior indicates that the barrier is low enough to have significantly populated local minima in the transition region, each with their own equilibrium constants that contribute to the unfolding curves measured. Finally, the most convincing evidence that  $\lambda$ nQ33Y is a downhill folder, is the fact that crowding agents do not affect reaction rates. Over the range of 130 to 160 g/ml Ficoll, no change was observed in the observed or folding rates measured, indicating that there is no barrier to lower.

## 4.5 References

1. Timasheff, S. N. (1993). The Control of Protein Stability and Association by Weak Interactions: How Do Solvents Affect These Properties? *Annual Review of Biophysics and Biomolecular Structure* **22**, 67-97.

2. Eggers, D. K. & Valentine, J. S. (2001). Crowding and Hydration Effects on Protein Conformation: A Study with Sol-Gel Encapsulated Proteins. *Journal of Molecular Biology* **314**, 911-922.
3. Zhou, H.-X., Rivas, G. & Minton, A. P. (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics* **37**, 375-397.
4. Ellis, R. J. M., A. P. (2003). Plugging the transport gap. *Nature* **425**, 27-28.
5. Qu, Y. X. & Bolen, D. W. (2002). Efficacy of macromolecular crowding in forcing proteins to fold. *Biophysical Chemistry* **101**, 155-165.
6. Ballew, R. M., Sabelko, J., Reiner, C. & Gruebele, M. (1996). A single-sweep, nanosecond time resolution laser temperature-jump apparatus. *Review of Scientific Instruments* **67**, 3694-3699.
7. Yang, W. Y. & Gruebele, M. (2004). Folding Lambda-Repressor at Its Speed Limit. *Biophysical Journal* **87**, 596-608.
8. Perry, R. H. & Chilton, C. H. (1973). *Perry's Chemical Engineers' Handbook*. McGraw-Hill Chemical Engineering Series, McGraw-Hill, Inc., New York.

## **Section III**

### **Bridging the Gap Between Scientific Research and K-12 Teaching**

#### **Introduction**

Whether by pushing for reform at the policy level or directly teaching in a K-12 classroom, research scientists can have an enormous impact on science education. Due to the early success of some innovative programs (see Appendix B.4), there has never been more support for such endeavors in terms of interest among graduate students, faculty and K-12 teachers, as well as ample funding at the state and national level.

From the perspective of K-12 institutions, there has never been a greater need for scientist participation in education reform as many sobering studies (Baldi et al., 2007) have exposed the growing science and math deficiencies among American children, especially compared to other industrialized nations. Future career opportunities for these children grow more uncertain, as fewer jobs are available domestically which do not require at least high school level math and science skills.

The actual task of K-12 science education reform is enormous and made worse by increasing numbers of students who show a below-proficient level of scientific understanding (Grigg et al., 2005). At the same time K-12 institutions are charged with preparing an increasing number of students for university level coursework and citizenship in an increasingly technologically complex society. With all of these additional pressures, teachers have limited time for innovation and continued professional development.

Research scientists can help in many ways: by organizing and distilling the wide range of emerging research in their field for a K-12 audience; by developing updated lesson materials and curricula; by leading in-class enrichment activities; and by providing a greater presence of scientists in K-12 classrooms.

One of the most important assets that scientists can bring to the K-12 classroom is an inquiry-based approach to learning. This is foreign to many K-12 teachers as much of the experimentation they use in their classrooms is meant to reaffirm what is already

known; therefore the idea of an activity that doesn't produce clear results or which leads to the conclusion that more research is necessary, may be hard for some teachers to accommodate. Scientists can help students and teachers learn to feel more comfortable saying "I don't know" and as a result seek to discover the answer.

University outreach programs also have the capacity to promote diversity in science by partnering in schools with low-income and under-represented populations. These students often have very little connection to institutions of higher education despite their close proximity in many cases and are also unlikely to know any working professionals in scientific fields. For these students, the simple presence of scientists in the classroom, particularly those who happen to be women or under-represented minorities, is already a great benefit. Outreach to these students can be an effective way to change their perceptions about who scientists are and who can become a scientist. Furthermore, whether or not these students pursue scientific careers, learning to think like scientists—by asking questions, making observations and deriving logical explanations—can be a source of great empowerment for disadvantaged students because these analytical skills are universally applicable to problems in every facet of life and learning.

Despite all of the important ways in which scientists can impact K-12 education, many shy away from such projects because of a lack of experience with younger learners or any training in pedagogy. Further, many scientists do not know how to get involved in outreach or may not even realize the enormous impact they can have. To address these problems, the following section presents two K-12 outreach projects completed as part of a science doctoral program in biophysics. These projects provided an excellent opportunity for experience working with K-12 schools and offered curriculum development, evaluation and teaching experience. Furthermore, these projects represent authentic forays into the formal discipline of science education, allowing a science graduate student the rare opportunity to make a significant contribution to the field.

## **References**

Grigg, W.S., Lauko, M.A., and Brockway, D.A. (2006). "The nation's report card: Science: 2005." Washington D.C.: NCES

Baldi, S., Ying, J., Skemer, M., Green, P.J., Herget, D., “Highlights from PISA 2006:  
Performance of U.S. 15-year-old students in science and mathematics literacy in  
an international context.” Washington D.C.: NCES



## Chapter 5

### **Research Tools in the Classroom: Integrating Molecular Visualization and Bioinformatics into the High School Biology Curriculum**

#### **5.1 Introduction**

Though it may be unpopular with teachers and students alike, a significant amount of memorization is necessary for students taking introductory science courses. This is in part because many new terms are indispensable for communicating science, which, like all disciplines, has its own vocabulary. In most cases though, teachers can reduce cognitive load by making connections to student's prior knowledge.<sup>1;2</sup> To make connections to well-known phenomena, teachers employ lectures, images, class demonstration and hands-on labs. In the molecular sciences, however, macroscopic observations and two-dimensional images are limited in their ability to convey all of the detail needed and so a formidable hurdle exists when traditional teaching methods are used. This fact was realized long ago in chemistry education where physical ball and stick models became a mainstay of the curricula long before the advent of computers. Over the past 20 years, computer visualization tools in chemistry have also become integral to introductory college and high school courses, a move which is supported by a wealth of education research.<sup>3;4;5;6</sup> The incorporation of physical and computer molecular visualizations into introductory biology curricula, however, has been slow especially at the pre-college level. This is despite the development in recent years of

powerful, user-friendly software and an increasing amount of lesson materials that are specifically geared toward high school learners. Whether the under-utilization of these resources by biology teachers is due to ineffective dissemination, fear of technology or some other factor is unknown. In fact, very little has been done to investigate the use of computer-aided molecular visualizations (CMVs) in high school biology instruction, and the educational merit of such activities has not been formally demonstrated.

To address these issues, we have developed and implemented five molecular visualization lessons employing the software “Visual Molecular Dynamics” (VMD),<sup>7</sup> in introductory, honors and AP biology courses at Danville High School (DHS) in Danville, Illinois and Neuqua Valley High School (NVHS) in Naperville, Illinois. We set out to determine whether CMV lessons could be made accessible to high school students and teachers, despite the limitations in class time and resources. We also set out to determine the educational impact of CMV lessons, especially considering that most introductory biology students have not yet had chemistry. Specifically, we wanted to know if CMVs could help students relate better to molecules, as measured by their understanding of intermolecular forces, molecular geometry and the size and scale of molecules and cells.

DHS was the primary laboratory for this study and all five lessons were implemented at this school. DHS is a good laboratory for assessing the feasibility and educational impact of CMVs since the academic challenges faced by this school are comparable to low-performing public schools nationwide. The school has consistently failed to meet Annual Yearly Progress (AYP) as the school’s approximately 600 Black students fell 20 points below AYP in reading and 27 points below in math, while the roughly 900 Caucasian students barely passed. The dropout rate is also quite high at

DHS with 8.4% of students not making it to graduation, versus a 3.4% average for the state. These facts are not unexpected as nearly half the student body is considered low income, living at or below the US poverty line (less than \$20,000 yearly income for a family of 4). Despite these statistics, DHS enjoys a well-educated faculty with most teachers possessing master's degrees. In addition, all math and science teachers have an undergraduate degree or other certification in their discipline. As a result, there are several Advanced Placement & dual credit courses, science cohorts known as "academies," and extracurricular science activities such as the Engineering Club, which won the "Class Projects" competition three years in a row at the University of Illinois' annual "Engineering Open House." Over the past few years, the science program at DHS has experienced an overhaul as educators attempt to push more students from the lower-level general science classes into college preparatory laboratory courses. This, combined with the limited number of spaces available in the school's honors program, leaves a range of learners in the introductory biology laboratory that demands highly differentiated instruction. For college-bound students, the pressure is particularly high since they must compete for coveted spots at the University of Illinois with students from Chicago's affluent suburbs, where public schools consistently rank among the top in the nation.

NVHS is one such nationally ranked, suburban Chicago school<sup>8</sup> and one of the CMV lessons developed here was implemented in NVHS honors biology classes. NVHS is more than twice the size of DHS, with 4500 students, of which less than 2% fall below the poverty line and only 4 or 5 drop out each year. The 3243 Caucasian students scored more than 15 points above the state minimum for AYP in both math and reading, while

the 626 Asian students scored more than 25 points above AYP in mathematics. The 300 Black students, however, scored 12 points below AYP in reading and 20 points below in math; however none of these students were represented in the honors biology classes in which these CMV lessons were implemented. As with DHS teachers, the majority of the NVHS faculty hold master's degrees and all science and math classes are taught by "highly qualified" teachers as defined by No Child Left Behind legislation.

DHS and NVHS biology classes represent a wide range of teachers, learners and learning environments that approximate the full spectrum of what can be expected in schools across the country. The physical, technical and time limitations as well as the educational impact of these interventions on the range of learners in this study will be widely applicable to high school biology classrooms across the country.

## **5.2 Materials and Methods**

### **5.2.1 Software**

VMD is freely available at [www.ks.uiuc.edu/vmd](http://www.ks.uiuc.edu/vmd). See Appendix C1.1 for download and installation instructions. The software requires 35 megabytes of disk space on a computer running Microsoft Windows 2000, XP, or Vista, or MacOSX 10.3.5 or later. Most lessons are easily adapted for use with another program such as Jmol or Pymol, since the structures used are all in pdb format which is readable by most molecular visualization software. The simulations must be converted into movies, however, before they can be used outside of VMD.

### **5.2.2 Classroom and Teachers**

#### *DHS*

Students worked in groups of two to four on desktop computers with Ethernet web access in a computer lab. The teacher's computer was connected to an LCD projector and a Smartboard screen. For lesson 1 and student presentations, the regular classroom was used, which has one computer connected to an LCD projector and Ethernet web access. Technical support staff were responsible for software installation. Space was provided for file storage and sharing on local DHS servers. Lessons 1, 2, 3 and 4 were implemented in one honors course and two introductory courses with teacher Kathryn Hafner during the 2007-2008 academic year. In addition, lesson 3 was implemented in one honors course, two introductory courses and one AP biology course with teacher Shelley Barker during the 2006-2007 academic year.

#### *NVHS*

Students worked in groups of three on a set of dedicated classroom laptops with wireless web access. An LCD projector and screen were used by the instructor for tutorials and demonstrations and again during student presentations. Technical support staff was responsible for loading the software. File sharing and storage were done using Blackboard on NVHS servers. Lesson 3 only was implemented in the five honors biology classes with teacher Matthew Kirkpatrick during academic years 2006-2007 and 2007-2008 and in the five honors biology classes with teacher Chris Meuller during academic year 2006-2007.

All the students and teachers in this study were novices with respect to VMD in particular and computer-based molecular visualization in general.

### **5.2.3 Lessons**

The lessons presented here are all designed as an advanced follow-up to the standard course material. They are not recommended as introductions to any topic as students may be overwhelmed by having to learn new content while also learning how to interpret molecular visualizations. Lesson plans, student handouts, PowerPoint presentations, associated molecule or simulation files and molecule/simulations files are available online at <http://gk12-uiuc.net/FINAL/downloads.shtml>. Lesson materials are also available in Appendix C.

#### **Lesson 1: “What’s So Special About Water?”**

*Prior knowledge required:*

This lesson is best suited as a follow-up to a lecture covering the bulk properties of water and hydrogen bonding. This is usually covered during the beginning of the “Chemistry of Life” unit in the 3<sup>rd</sup> week of class at DHS. Prior to this lesson, students should be introduced to hydrogen bonds. Specifically, they should understand how hydrogen bonds give rise to most of water’s life-sustaining properties: its high surface tension, heat capacity, enthalpy of vaporization, melting and boiling points, the expansion of water upon freezing, and water’s ability to act as a “universal solvent.” Students should understand why the above properties are important for life on earth.

*Description of lesson:*

This is a one-day lesson involving a lecture and classroom discussion focused around two molecular dynamics simulations, one showing a cube of water in vacuum equilibrating to a sphere to minimize surface tension and a second showing an ice crystal melting. This lesson requires one computer with VMD installed, an LCD projector and a screen. Students do not interact directly with the software in this exercise. The simulations are presented by the instructor over an LCD projector. This exercise is expected to be a good first exposure to molecular visualizations, since students are led from flat two-dimensional images of water molecules, to 3-dimensional representations of water and ice crystals.

**Lesson 2: “Visualizing Biopolymers and Their Building Blocks”**

*Prior knowledge required:*

This lesson should be used as a follow-up to the biochemistry and molecular biology unit. At DHS, this material is covered in the “Chemistry of Life” unit between the 4<sup>th</sup> and 6<sup>th</sup> week of class. The students should have a basic understanding of polymers. Specifically, they should understand that:

1. Cells make a huge number of large molecules called polymers from a limited set of small molecules called monomers. When the monomers are identical, as with starch, we call it a homo-polymer. If there are two or more different monomer types, as with proteins, it is a hetero-polymer.

2. Polymers are formed in a condensation reaction (dehydration synthesis) in which two monomers lose a water molecule and are joined covalently. Hydrolysis is when the covalent bonds linking two monomers are broken and water is gained.
3. Biopolymers fall into four distinct classes: proteins, nucleic acids, lipids, and carbohydrates which are made up of amino acids, nucleosides, fatty acids and monosaccharides, respectively.

*Description of lesson:*

This is a one to four day lesson in which students use VMD to explore each of the four classes of biopolymers: lipids, proteins, carbohydrates, and nucleic acids. For each class, a worksheet was designed to lead students through an exploration of the molecular structure of monomers and polymers. Depending on the time available, one or more of the four sections may be skipped. This lesson requires a computer with VMD installed for the instructor and students. In addition, the instructor's computer must be connected to an LCD projector so he or she can lead students in a brief tutorial beforehand. Students may work independently or in groups of one to three. Ball and stick models can be used together with VMD if they are available.

**Lesson 3: “My Favorite Protein”**

*Prior knowledge required:*

This lesson is designed as a follow up to the molecular biology unit covered in weeks 12-14 at DHS. Students should understand the following prior to beginning this lesson:



1. Proteins are hetero-polymers of amino acids. There are 20 unique amino acids, some positively charged, some negatively charged, some non-polar, and some polar.
2. A protein's sequence refers to the exact type and order of amino acids. A protein's structure refers to its 3-dimensional form when properly folded.
3. A protein will fold up into a unique, compact structure, depending on what types of amino acids compose it and the order in which these amino acids are arranged.
4. Hydrophobic amino acids are non-polar, oil-like molecules. They will tend to associate with one-another when dissolved in water. This separation of oil-like and water-like parts of a protein is part of what makes it fold. Hydrophobic amino acids will be found on the inside of a folded protein, unless it is a membrane protein, which will have hydrophobic amino acids on the outside.
5. Other intermolecular forces important for protein stability include hydrogen bonds, ionic bonds (salt bridges), and covalent sulfur-sulfur (disulfide) bonds.
6. Enzymes are proteins that speed up (catalyze) reactions. They all have little pockets inside them called active sites where the reaction takes place.
7. Protein function is usually very sensitive to structural changes, particularly near an active site or binding site.

*Description of lesson:*

This is a challenging, three-day visualization lesson in which students use the VMD software to both explore a protein structure and synthesize their own 3-dimensional model. They will choose a protein from the Protein Data Bank's (PDB) "Molecule of the

Month,” accessible for free at [www.rcsb.org](http://www.rcsb.org). This lesson was written with a focus on enzymes, and students are reminded of the many enzymes, such as DNA and RNA polymerase, ATP synthase and Photosystems I and II that they have already encountered in class. This lesson can be easily adapted, however, to focus on membrane proteins, light-reactive proteins or another category that best fits the curriculum. After students choose a protein, they can download the information packet available from the PDB website to learn about their protein’s function and structure. They then load the X-ray crystal structure coordinate file (the PDB file) into VMD and use the program to highlight the parts of the protein that are important for its function. As a final step, students build a 3-dimensional model of their protein using common household materials. The models are meant to approximate the general shape of the protein and to highlight the details of the active site or other regions important for maintaining the protein’s stability and function. The lesson product is a written report describing the protein’s structure and function along with both a VMD image and a 3-dimensional model of the protein. Students are given one week to prepare their reports and physical models.

#### **Lesson 4: “Name That Gene, Disease & Protein”**

##### *Prior knowledge required:*

This lesson is designed as a follow up to the genetics unit covered in weeks 2-4 of the second semester at DHS. Students should understand the central dogma of molecular biology and have a working knowledge of the terms gene, nucleotide, base, base-pair, centromere, chromosome, genome, mutation, genetic disease, genotype, phenotype, bioinformatics, sequence alignment, BLAST and query. In addition, students should

understand that a mutation can be a small change in just one nucleotide or the change can involve many nucleotides and that the normal nucleotides can be substituted for something else, deleted altogether, or an extra nucleotide can be added.

*Description of lesson:*

This lesson is adapted from Wefer et al<sup>9</sup> and adds a molecular visualization component to the standard “Name That Gene” activity in which students are required to use BLAST searches and the National Center for Biotechnology Information (NCBI) “Genes and Disease” online book to determine the genetic disease associated with a small fragment of DNA they are given in class. The molecular visualization component requires students to go one step further and highlight the disease causing mutation in the protein structure. Based on visual inspection of the structure and information from the “Genes and Disease” online book, students must determine how the mutation changes the function or stability of the protein to produce the observed pathology. The lesson product is a report that explains the identity of their gene, their gene’s location in the genome, the symptoms of the genetic disease associated with their gene and an explanation of how the protein’s function is altered by the disease-causing mutation(s).

**Lesson 5: “Proteins as Molecular Clocks”**

*Prior knowledge required:*

Students should be very familiar with the hierarchy of biological classification and should be comfortable with the terms phylogenetics, cladistics, taxonomy, convergent evolution, horizontal gene transfer, and genetic drift. This should not be their

first time seeing evolutionary relationships displayed in cladogram/tree form. It is highly recommended that this activity be preceded by a pen and paper activity using limb morphology or short DNA/protein sequences to build evolutionary trees.

*Description of lesson:*

This 1-2 day lesson is designed as a follow up to the evolution unit, which is covered in the last month of the semester at DHS. In this lesson, students explore molecular evidence for evolution at the level of proteins. First, the students choose a group of organisms to classify and three proteins to use for the comparison. The group of organisms must be closely related (i.e. insects, flowering plants, carnivores, etc.) and the three proteins must be present in all the organisms (i.e. hemoglobin in higher animals, cytochrome C in bacteria, etc.). Students use Multiseq, a bioinformatics plugin in VMD, to perform a BLAST search of multiple genome databases for sequences of their desired protein. They can then filter according to the different organisms they want to classify. They then perform multiple sequence alignments of each protein and use them to build phylogenetic trees. Using the Multiseq plugin, students can easily see where sequence differences arise and click on those amino acids in the sequence alignment to highlight them in the VMD protein structure. Students are asked to inspect the protein structure to determine where most of the sequence variation occurs and explain their observations. Students are also asked to compare the phylogenetic trees they created to the established evolutionary relationships (available online from <http://tolweb.org/tree>) and to explain how discrepancies can arise in the trees created for the 3 different proteins (e.g. due to convergent evolution or lack of sufficient changes among a very closely related group).

For a more advanced lesson, this activity can be extended to include multiple structural alignments of proteins in addition to sequence alignments. Sequence and structure trees can be compared to determine which alignment method better approximates established taxonomic relationships and which are more internally consistent.

### **5.2.4 Assessment**

Qualitative and quantitative methods were used to address the following research questions:

1. Is it feasible to use computer molecular visualizations (CMVs) in high school biology? Can we run the software on existing resources? What are the limitations or potential obstacles? Can teachers and students load and manipulate molecules and color them in meaningful ways in VMD? To what extent do students understand the molecule representations in VMD? Do they see only a jumble of lines and balls, or can they see atoms, covalent bonds, molecular geometry, overall shape, and intermolecular interactions? Can they distinguish molecules from one another as in, say, the water DNA, RNA and protein in the transcribing RNA polymerase structure? Can they make connections between molecule representations in VMD and textbook cartoons, chemical formulas, physical ball and stick models, line-angle, or Lewis drawings?
2. Do CMV activities help students understand molecular properties and how they give rise to macroscopic phenomena? Specifically, do they improve students'

- understanding of hydrogen bonding, hydrophobic interactions and the relative size and scale of atoms, molecules and cells?
3. Do CMVs increase the amount of inquiry learning in biochemistry and molecular biology units of the introductory biology course?

The primary assessment instruments used were surveys administered to DHS students. There are two different surveys, one post-activity survey (figure 5.1) administered to the 2006-2007 cohort and another survey (figure 5.2) administered to the 2007-2008 cohort at the beginning of the semester prior to any CMV exposure, and then again after the last CMV lesson was implemented. In addition to the surveys, teacher journals, videotapes of the “My Favorite Protein” classroom activity, student reports and models from the “My Favorite Protein” lesson, and student grades were all used to track obstacles to implementation and learning impact. Table 5.1 summarizes the instruments used for this study along with the data analysis methods employed.

**Table 5.1** Assessment methods.

<b>Instrument</b>	<b>Data Source(s)</b>	<b>Analysis Methods</b>
Journals	Teachers, Researcher	Track teacher and researcher views on costs and benefits of CMV lessons along with obstacles to implementation.
2007-2008 Survey	Students	Tally self-reported results of students' attitudes and learning.  Tally results to determine conceptual knowledge of proteins and relative size of molecules and cells. Tally results for each question and size pair separately as well as a combined "concept knowledge" score.
2006-2007 Survey	Students	Tally results to determine conceptual knowledge of proteins and relative size of molecules and cells. Tally results for each question and size pair separately as well as a combined "concept knowledge" score.  Compare pre and post survey results.
Classroom Video	Students, Teachers, Classroom Environment	Track students' apparent involvement/engagement  Track problems with implementation.
Lesson 3 Presentations, Reports and Models	Students	Describe the overall quality and level of thought employed, common mistakes, and the conceptual understanding required for synthesis of observed results.
Grades	Students	Compare semester grades with "My Favorite Protein" lesson grades.

**Figure 5.1** 2006-2007 Survey (continued next page).

- 1). Please describe the “My Favorite Protein” assignment in your own words. What did you do and what was the goal of the assignment?
- 2). Did you like this assignment? Rate this assignment from 1 to 5 according to how much you liked it. A rating of 1 means you really hated it and 5 means you loved it.
- 3). What was the **most** interesting part of this assignment (choose as many as apply)?
- \_\_\_\_\_ Using VMD.
- \_\_\_\_\_ Preparing and giving your presentation.
- \_\_\_\_\_ Learning about your protein’s structure.
- \_\_\_\_\_ Learning about your protein’s function.
- \_\_\_\_\_ Making your physical model.
- \_\_\_\_\_ Researching or writing your report.
- \_\_\_\_\_ other: \_\_\_\_\_
- 4). What was the **least** interesting part of this assignment (choose as many as apply)?
- \_\_\_\_\_ Using VMD.
- \_\_\_\_\_ Preparing and giving your presentation.
- \_\_\_\_\_ Learning about your protein’s structure.
- \_\_\_\_\_ Learning about your protein’s function.
- \_\_\_\_\_ Making your physical model.
- \_\_\_\_\_ Researching or writing your report.
- \_\_\_\_\_ other: \_\_\_\_\_
- 5). How much did you learn during this assignment?
- \_\_\_\_\_ I learned less from this assignment than I did from other assignments in the Organic Compound unit.
- \_\_\_\_\_ I learned as much from this assignment as I did from other assignments in the Organic Compound unit.
- \_\_\_\_\_ I learned more from this assignment than from the other assignments in the Organic Compound unit.
- 6). Please rate the difficulty of each part of this assignment from 1 to 5, where 1 means it was very easy and 5 means it was one of the hardest things you’ve done in this course.
- \_\_\_\_\_ Learning how to use VMD (the molecular visualization software)
- \_\_\_\_\_ Learning about protein structure and what was unique about your protein’s structure
- \_\_\_\_\_ Learning about the function of your protein
- \_\_\_\_\_ Writing your report
- \_\_\_\_\_ Preparing and delivering your presentation
- \_\_\_\_\_ Building your physical model



**Figure 5.1, continued** 2006-2007 Survey.

7). Please place the following in order of their size, from smallest to largest (I have filled in the largest and smallest for you already):

- |                       |  |
|-----------------------|--|
| a) a protein          | Smallest: 1) <u>e. a hydrogen atom</u> |
| b) a mitochondrion    | 2) _____                               |
| c) a cell             | 3) _____                               |
| d) a ribosome         | 4) _____                               |
| e) a hydrogen atom    | 5) _____                               |
| f) a grain of sand    | 6) _____                               |
| g) one water molecule | Largest: 7) <u>f. a grain of sand</u>  |

8). The shape of which of the following is most similar that of most proteins? (choose one)

- a) A straight, thin wire      b) A flat compact disc (CD)    c) A round ball

9). Which of the following statements is true about most proteins (choose one):

\_\_\_\_\_ Most proteins are static, they don't move at all.

\_\_\_\_\_ Most proteins can move around in space (diffuse) as a part of normal function and behavior, but cannot vibrate, flex, or change their shape.

\_\_\_\_\_ Most proteins vibrate, flex, and change their shape as a part of normal function and behavior, but they do not move around in space (diffuse).

\_\_\_\_\_ Most proteins move around in space (diffuse), vibrate, flex, and change their shape as a part of normal function and behavior.

10). Which of the following are responsible for holding the 3-dimensional configuration of a protein together? (circle all that apply)

- a). Peptide bonds      b). Hydrogen bonds      c). Disulfide Bonds      d). Hydrophobic core

11). When studying Biology, we sometimes need to memorize facts and we sometimes need to do independent research or problem solving. What ratio of memorization to independent research and problem solving did you do during this assignment? (choose one)

\_\_\_\_\_ I spent most of my time on memorization of facts for this assignment.

\_\_\_\_\_ I spent about half of my time on memorization of facts and about half on independent research or problem solving.

\_\_\_\_\_ I spent most of my time on independent research &/or problem solving for this assignment.

12). What does the protein you studied do? Explain your protein's **function** as though you were speaking to a 5<sup>th</sup> grader.

13). What does your protein look like? Explain your protein's **structure** as though you were speaking to a 5<sup>th</sup> grader.

**Figure 5.2** 2007-2008 Survey.

**Instructions:** Please answer the following questions to the best of your ability. The questions may seem difficult, but do your best to think about what you already know about the subjects being discussed and then use this information to come up with your best guess.

1). Please place the following in order of their size, from smallest to largest:

- |                           |                    |
|---------------------------|--------------------|
| h) a chromosome           | Smallest: 1) _____ |
| i) a gene                 | 2) _____           |
| j) a malaria cell         | 3) _____           |
| k) a human red blood cell | 4) _____           |
| l) a hydrogen atom        | 5) _____           |
| m) a mosquito             | 6) _____           |
| n) a proton               | 7) _____           |
| o) a protein              | 8) _____           |
| p) a water molecule       | Largest: 9) _____  |

2. Please draw a picture that shows how all the above items are related. Please make sure to label the items in your drawing.

3). What does a typical protein look like? Please describe its appearance in everyday language.

4). Please answer the following True/False questions. Explain each answer by writing what you know about atoms, cells, and molecules that helped you decide on your answer.

DNA is made up of cells \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

Atoms are made up of cells \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

Cells are made up of atoms \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

Cells are made up of molecules \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

Molecules are made up of atoms \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

Everything is made of cells \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

Everything is made of atoms \_\_\_\_\_ True \_\_\_\_\_ False  
Explain:

5). Describe the qualities you would have if you were a water molecule? How would you look, behave, & interact with others? What would you do inside a cell? Be as detailed as you can (you may also draw a picture).

## **5.3 Results**

### **5.3.1 Journals**

Table 5.2 shows the most salient comments from teacher and researcher journals for each of the 5 VMD lessons. Both teacher and researcher's journal comments reveal that the primary obstacles to VMD lesson implementation were restricted access to school computers and an overburdened technical staff unfamiliar with this type of software. There were several instances when the technicians installed the software, but did not make it accessible to students. File sharing between students and teachers was also a recurring problem that interfered with several lessons.

Overall, teachers thought the activities were more worthwhile for advanced students, while it was unclear what students in the introductory classes were able to learn, since they were absent often, making it difficult to properly prepare them for these advanced lessons.

### **5.3.2 Surveys for 2006-2007 Cohort**

Survey responses were analyzed for 22 introductory biology, 29 honors, and 23 AP biology students from DHS, approximately one month after the “My Favorite Protein” activity. We analyzed students' self-reported attitudes and their conceptual knowledge. Figure 5.3 shows that most students enjoyed the lesson and felt that building both VMD and physical models were relatively easy. Only AP students completed physical models due to limited resources and the request by the teacher to simplify the

assignment for less advanced students. The most difficulty students reported having, was learning the

**Table 5.2.** Summary of journal comments (continued next page).

<b>Obstacles</b>	<b>Costs</b>	<b>Benefits</b>
VMD not installed on time on classroom computer	No hands-on component, mostly passive learning	“Showing” rather than “telling” about hydrogen bonding
VMD not installed on time in computer lab computers  Was too long/challenging to complete in allotted time (2 class periods)	Not clear how this activity is superior to those using ball and stick physical models	Allowed students to “see” molecules.  Problem solving rather than memorization.
DHS servers would not allow students to save to common directory that teachers can access.  “Web PDB Download” feature of VMD was extremely slow due to limited bandwidth and simultaneous requests of entire class.	A lot of information over a brief period of time  Printing of VMD images in color was difficult (not available at school) and expensive.  Teacher must read up on each genetic disease, protein, mutation, etc. in order lead lesson and grade reports.	Challenging activity for advanced students  Less advanced students remain engaged due to entertainment value of software and creative aspects of the assignment  Exposes students to research tools used by real scientists.

**Table 5.2, continued** Summary of journal comments.

Obstacles	Costs	Benefits
<p>“Web PDB Download” feature of VMD was extremely slow due to limited bandwidth and simultaneous requests of entire class.</p>	<p>Limited time (2 days) was not sufficient.</p> <p>Many concepts in one lesson – overwhelming for some students.</p> <p>Teacher must read up on each genetic disease, protein, mutation, etc. in order lead lesson and grade reports.</p>	<p>Exposes students to research tools used by real scientists.</p> <p>Helps students see the link between genotype and phenotype and reinforce learning of central dogma.</p> <p>Exciting &amp; personal connection to genetics for students.</p>
<p>Problem with Multiseq metadata file saving, since saving to hard drive is not allowed.</p> <p>Properly configuring Multiseq was too difficult for technical staff (had to be done by teacher/researcher)</p>	<p>Lesson is too advanced for general biology students</p> <p>Teacher must read up on each protein and taxonomic group in order to grade reports.</p> <p>Many proteins don’t have PDB structures or the sequences aren’t available for chosen organisms.</p>	<p>Lesson challenges AP and other advanced students</p> <p>A new and interesting way to learn evolution and genetics concepts.</p>

important features of their protein’s structure. Many student models and reports corroborated this result.

Figure 5.4 shows what students know about the relative size of atoms, molecules and cells. The results point to a solid knowledge of cell biology on the part of most students. Atoms and molecules are much more abstract for students. About half of students thought that water molecules were larger than proteins, ribosomes,

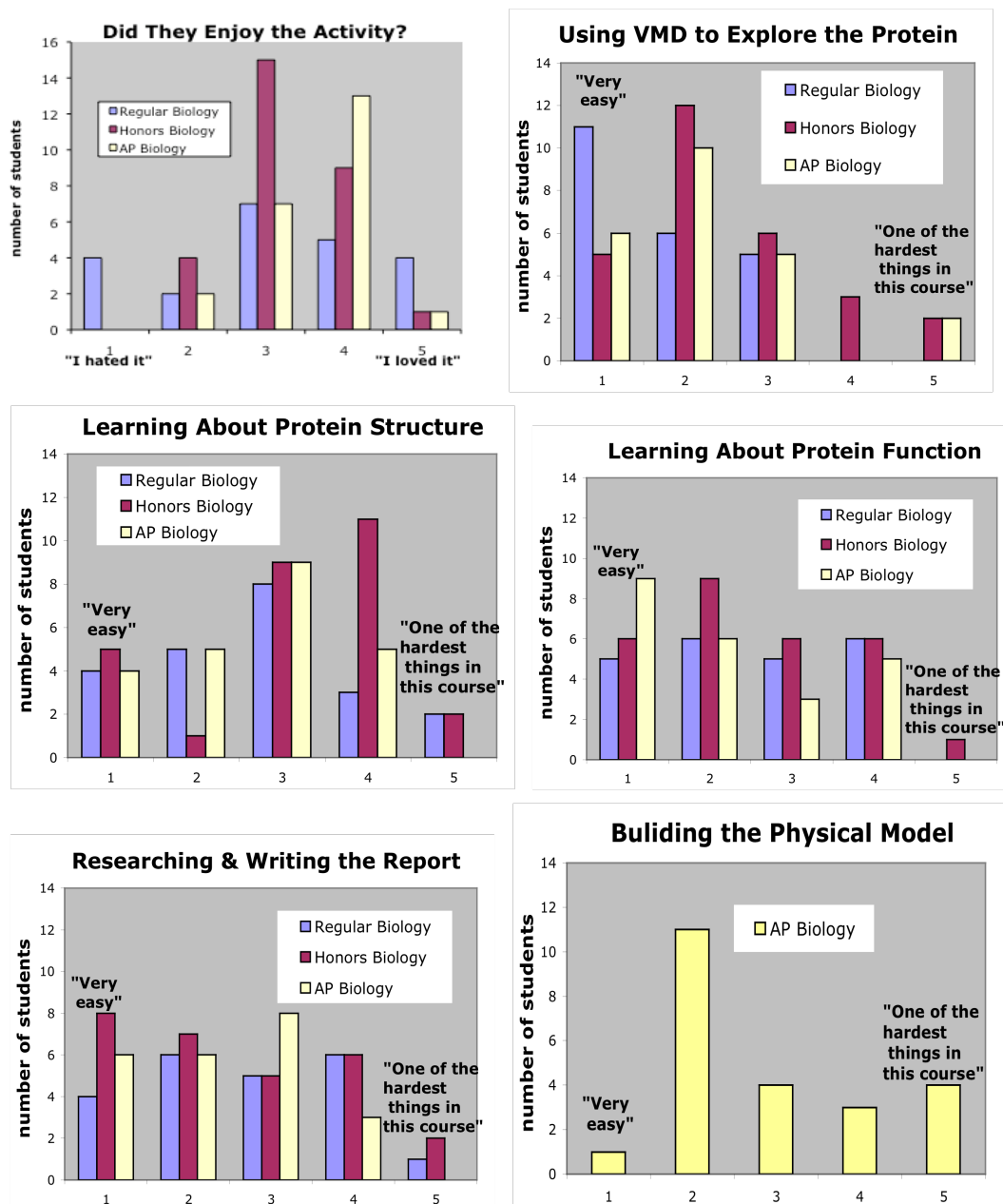
mitochondria, and even cells. The most incorrect answers were received for the ribosome-protein relative size, though this was likely confusing for students since the ribosome is itself a protein complex. The question was meant to rank the size of an average protein, which would be much smaller, but this was unclear in the survey wording.

Figure 5.5 shows results for survey questions that tested students' knowledge of bonds important for protein stability. Although considerable class time was devoted to hydrogen bonds, only a few introductory biology students thought they played an important role in stabilizing proteins. The same was true for peptide bonds, although the "Organic Molecules" unit covers the basic chemistry of peptide synthesis. Disulfide bonds and hydrophobic interactions were the bonds most students were able to recall as being important in proteins. This is possibly because this type of bonding was introduced during the protein activity and was not discussed in other contexts in the course. Though this survey was given a month after students completed the activity, about half of students could still explain the function of their chosen protein from the "My Favorite Protein" lesson. Only 10 AP students and 2 honors students could still remember and explain the details of their protein's structure, however. Most students did not even complete this question or wrote "I don't remember".

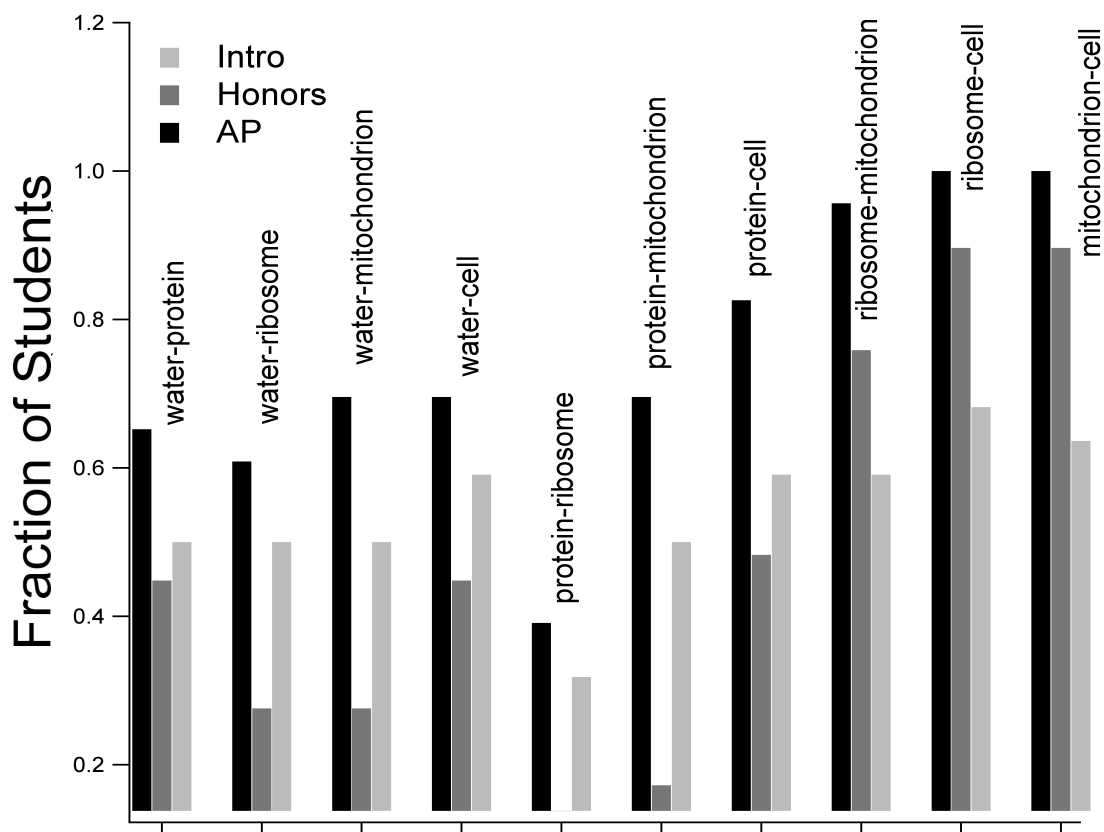
Figure 5.6 shows results for student learning, both self-reported learning and learning measured by basic knowledge survey questions and student reports/presentations. Less advanced students felt that they learned more than more advanced students, even though they scored far lower, on average, with both the "My Favorite Protein" reports, presentations and models and on the basic knowledge survey

questions. Advanced students also appeared to have a better grasp of the assignment goals and expectations, compared with introductory and honors students.

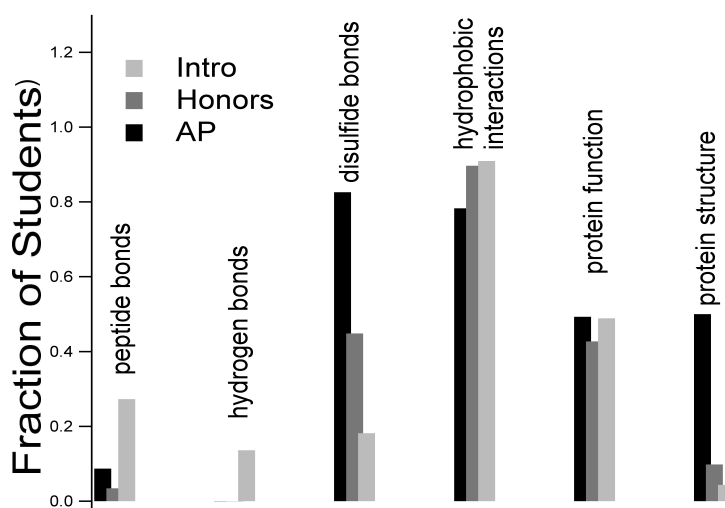
**Figure 5.3** DHS student's attitudes about "My Favorite Protein" activity (2006-2007).



**Figure 5.4** 2006-2007 DHS student's knowledge of relative size of molecules and cells.

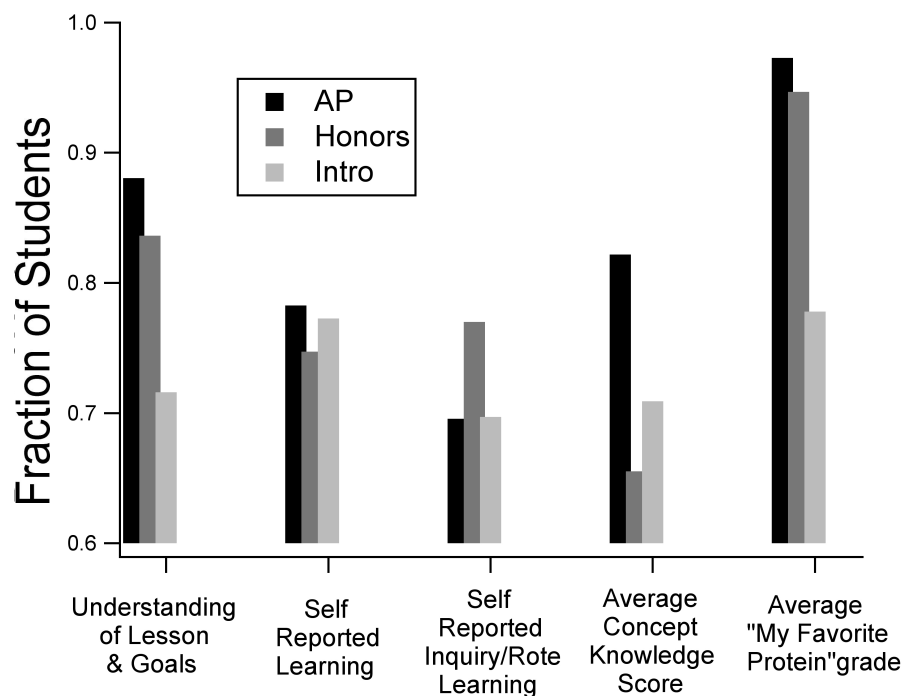


**Figure 5.5** 2006-2007 DHS Student's understanding of proteins.





**Figure 5.6** DHS Student Learning (2006-2007).



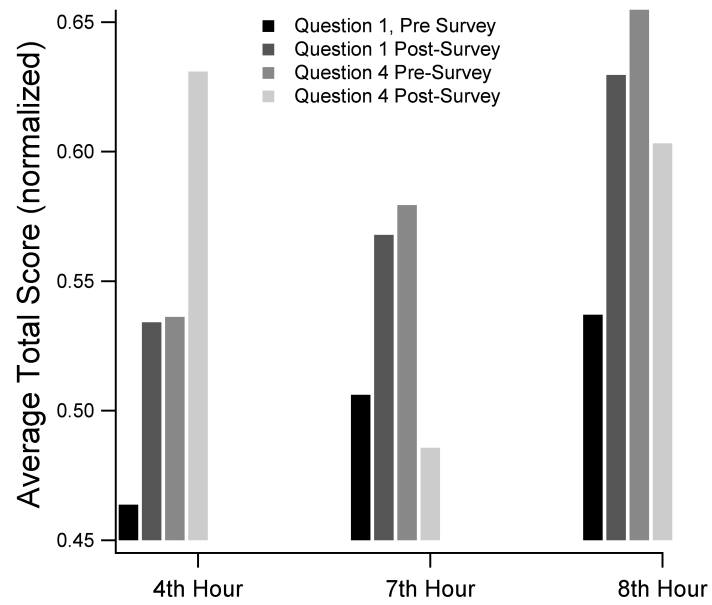
### 5.3.3 Surveys for 2007-2008 Cohort

Substantial changes were made to the survey for this cohort. We kept the size-ordering question (changing only a few items in the list) and we added another question (question 4) to test students' understanding of the composition of cells and molecules. In addition, we added three open-ended questions to further probe student thinking. Our goal was to elicit responses that might provide details about how students view molecules and cells. These questions were abstract, requiring a significant amount of critical thinking, asking students to make connections between previously isolated concepts, each of which was (at best) vaguely understood. The unfortunate result was that students generally left the open-ended question blank. In addition, even with the true/false (question 4) and size-ordering (question 1) questions, there was an indication that

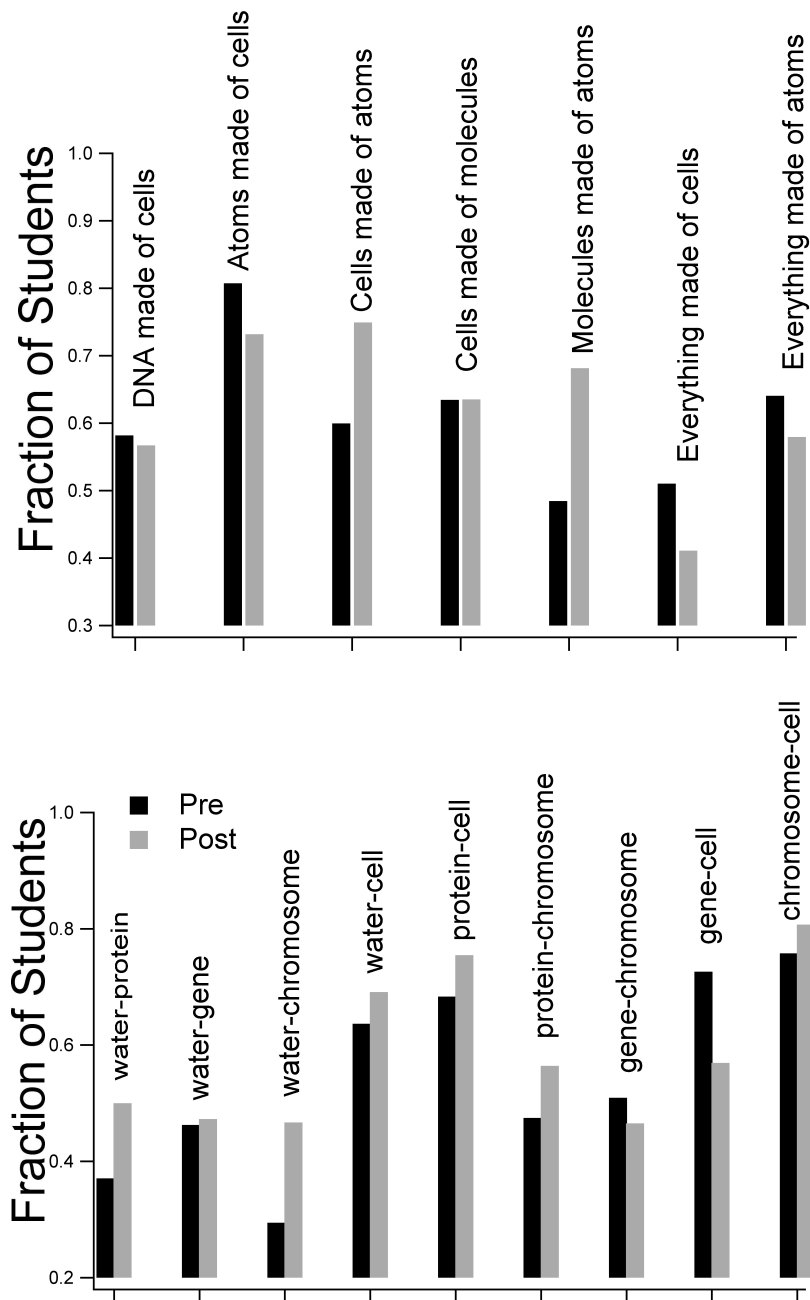
students did not put in significant effort or thought. Very few students wrote an explanation for any of the true/false statements in question 4. The few that did write anything simply restated the question. In addition, several students answered “true” to both the “atoms are made of cells” and “cells are made of atoms” sections of question 4. Almost every student who answered “true” for “cells are made of atoms” and “molecules are made of atoms” answered “false” for “cells are made of molecules.” The increased difficulty of this survey, compared with the one used in 2006-2007, should have been accompanied by increased time for completion; however, time constraints from a demanding curriculum and several missed classes due to snow and holidays meant that only a brief time-window (20 minutes) was available for completion of the survey.

Figure 5.7 shows the average students’ scores for questions 1 and 4 before and after the CMV lessons. All classes showed improvement on the size ordering in question 1, but only 4<sup>th</sup> hour students showed improvement for question 4. Figure 5.8 shows the breakdown of scores for each subsection of question 4 and selected size pairings from question 1. The top panel shows that the only improvement after the CMV lessons was that more students reported that both cells and molecules were made of atoms. There was marginal improvement in the size pairing activity in every category except gene-cell and gene-chromosome, though the class finished both genetics and cell biology units before completing the second survey.

**Figure 5.7** Total Score for Questions 1 and 4.



**Figure 5.8** Total score for each subsection of question 4 (top panel) and selected size pairs from question 1 (bottom panel) for the 2007-2008 cohort.



### 5.3.4 Video

Due to technical problems, over half of the video contains no sound. The visual-only and audible portions, however, still reveal several interesting details about the “My Favorite Protein” lesson.

1. Most students rely heavily on help from teachers and appear to have difficulty following instructions from the handout or problem solving on their own.
2. The above problems seem to be alleviated somewhat in “effective” groups where each member is actively involved and communicating often (about the project) with fellow group members (figure 5.9a).
3. Students seem to enjoy playing with the software and make creative images, but it is unclear if they understand what they are looking at.
4. Low performing students appear more engaged, but it is unclear what they are learning.
5. The activity is demanding for teachers who must attend to many simultaneous questions by groups studying different proteins (figure 5.9b).
6. Apparent frustration is observed when students raise their hand, but teachers are helping someone else. This happens quickly (within 5 minutes).
7. The class period is short compared to material/concepts covered and tasks required.

**Figure 5.9** “My Favorite Protein” classroom video. Students were most engaged and productive when there was good communication among group members (top panel) or with an instructor (bottom panel).



### 5.3.5 “My Favorite Protein” Presentations, Reports and Models.

Examples of student work are shown in figures 10-12. Figure 5.10 shows six of ten slides from the Power Point presentation of an NVHS biology student group that chose to study the protein dihydrofolate reductase. This presentation represents the best of what was observed across both schools and both 2006-2007 and 2007-2008 cohorts. Students clearly describe the protein’s structure and function in their own words and provide several VMD images to illustrate these descriptions. The description of the protein’s structure as “tertiary” refers to the fact that it is a single chain. This points to some confusion on the part of even the most advanced students with the hierarchy of protein structure and associated terms (primary, secondary, tertiary, quaternary). The physical model for this group is shown in figure 5.11a, and also indicates a solid understanding of the assignment and of the protein’s structure. The model approximates the overall shape, points out disulfide bridges in yellow, and attempts to model the molecular detail of the active site.

The model in figure 5.11b also does a good job of modeling overall shape while detailing secondary structure elements and producing a highly detailed and accurate molecular model of the Luciferin cofactor. The ATP synthase physical model in figure 5.11c reveals a distorted understanding of the size of the F1 transmembrane proton pump relative to the soluble F0 synthase domain. This may be due to the fact that the PDB structure contains only the synthase domain, so that any information about the remainder of the protein structure must have been obtained from images – possibly cartoons from their textbook.

Figure 5.12 shows VMD and physical models for three student groups from DHS. Despite the lower level of DHS students compared with NVHS students, many students were able to create meaningful models that reveal the unique features of their protein's structure and function. The best example is shown in figure 5.12b, the model for Cholera Toxin. Though the VMD drawing & coloring scheme is very simple, the group's physical model and demonstration during their presentation revealed an excellent understanding of the protein's structure and function. This group held up the model and explained that the yellow part sticks in the membrane, while the alpha helix (red) delivers the toxin (blue) by shoving it through the hollow center of the yellow part. The students demonstrated this action with their model, indicating a clear understanding that the protein's structure is uniquely adapted for its function, and that the protein is a dynamic structure that changes its conformation to accomplish a task.

Figure 5.12c shows the VMD and physical model for a group studying DNA Polymerase. These structures are representative of misconceptions on the part of close to half of DHS students. First, these students did seem to understand the assigned task, namely to approximate the general shape while detailing the active site, but they were clearly confused by the VMD program. This group could not come up with a drawing method and coloring scheme to highlight the nucleic acid in the active site, while choosing a more general drawing method for the overall protein structure. They were able to create a new representation for the DNA and color it by "Chain," and drawing method "VDW." This would have yielded an acceptable and informative image if they had either left the protein representation with the drawing method "Lines" or if they had changed it to "New Cartoon" or "Tube." Instead, they chose the same drawing method



for the protein as they did for the DNA (VDW) and they kept the coloring method (Name) making it very difficult to see the protein's active site and the DNA inside. The resulting VMD image is a jumble of differently colored balls that give no indication of how the protein functions. The physical model is also troubling, in that it is 2-dimensional. This may be because the group prepared their model at home with only this VMD image as a guide.

Figure 5.12a shows the VMD and physical model for the DHS student group that chose Myoglobin. The VMD model in this case is the scanned image from one student's report. The first two pages of the report are shown in figure 5.13. The student's name is covered in both figures. This is another example of a DHS student who mastered the concepts in this lesson. The student explains in detail both their physical and VMD model, labeling the heme, the iron and the alpha helices. He incorrectly refers to a bound oxygen molecule as a water, but he demonstrates an awareness of the CPK coloring scheme and may have been confused by the fact that the waters in the structure do not have hydrogens. Unfortunately, many of the student reports contained plagiarized sections and/or descriptions that made little sense. This student does a good job, however, of describing the structure and relating it back to protein function. They also do a good job of explaining the importance of myoglobin for mammals and describe its importance in the scientific community as the first protein to have its high resolution structure elucidated by X-ray crystallography.

**Figure 5.10** NVHS Student Power Point Presentation.

### About Dihydrofolate Reductase

#### Function

- This protein plays a supporting role in the building of DNA and other processes.
- It manages the state of folate.
- Folate is a snaky organic molecule that shuttles carbon atoms to enzymes that need them in their reactions.

### About Dihydrofolate Reductase

#### How does Dihydrofolate Reductase achieve its function?

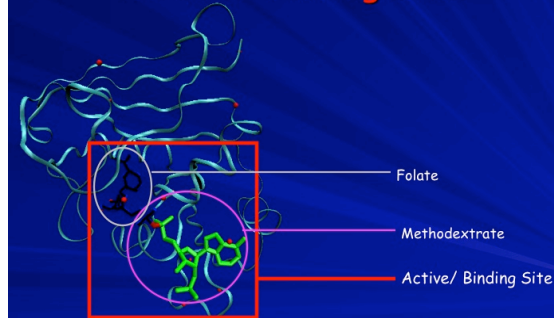
- Dihydrofolate Reductase has a long groove that binds to folate at one end and to NADPH at the other end.
- NADPH provides the hydrogen atoms to be transferred to the folate molecules by the Dihydrofolate Reductase.
- Then, the enzyme transfers hydrogen atoms from NADPH to the folate, converting folate to a useful reduced form.

### About Dihydrofolate Reductase

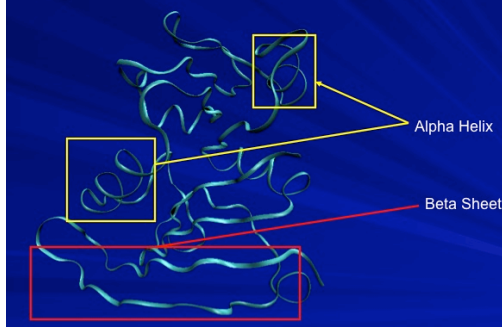
#### How does Dihydrofolate Reductase achieve its function?

- Dihydrofolate Reductase has a long groove that binds to folate at one end and to NADPH at the other end.
- NADPH provides the hydrogen atoms to be transferred to the folate molecules by the Dihydrofolate Reductase.
- Then, the enzyme transfers hydrogen atoms from NADPH to the folate, converting folate to a useful reduced form.

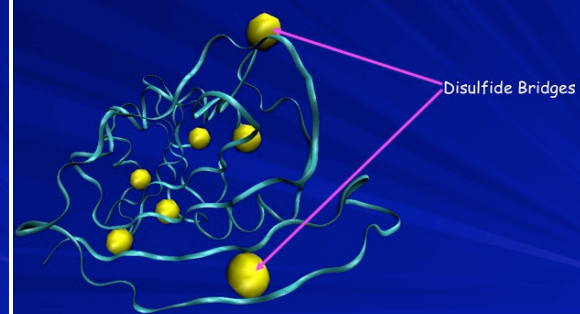
### Active/ Binding Site



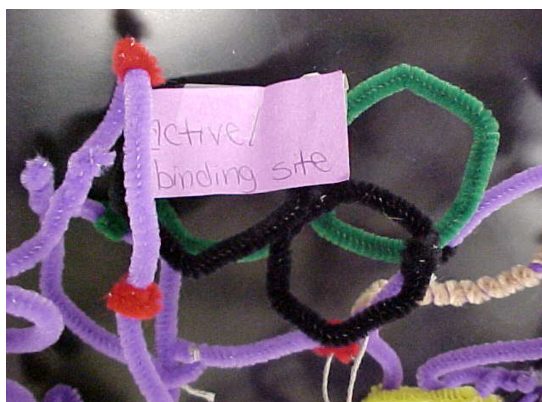
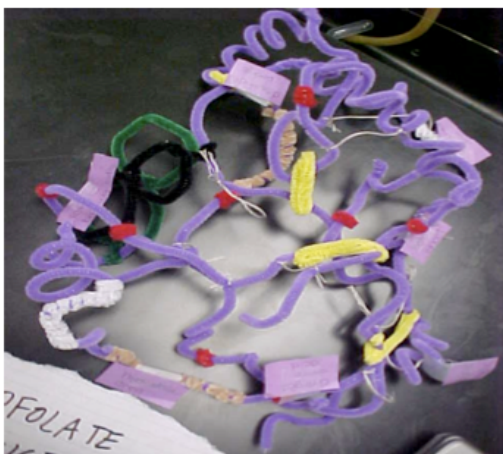
### Alpha Helices and Beta Sheets



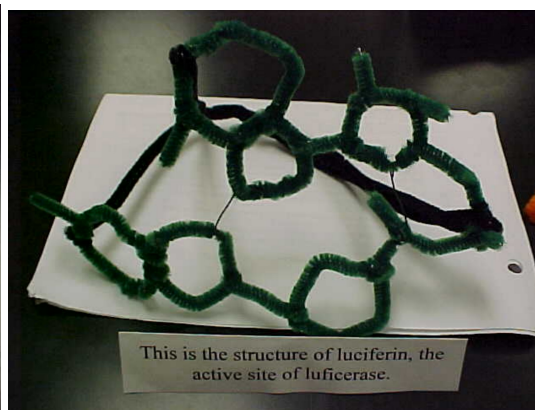
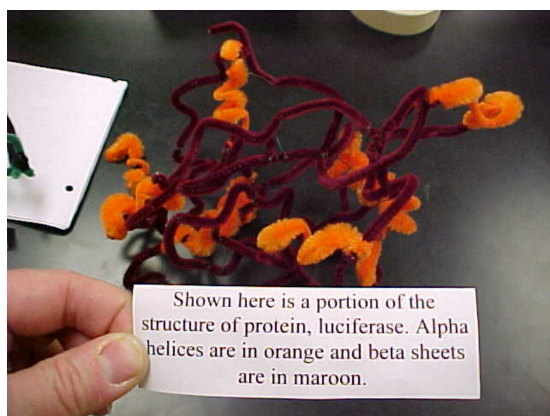
### Disulfide Bridges



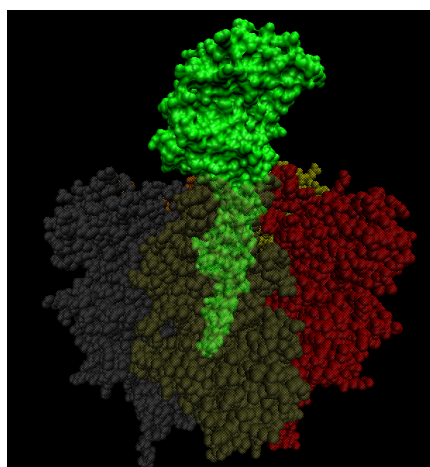
**Figure 5.11** NVHS Students' VMD and Physical Models.



### A. Dihydrofolate Reductase



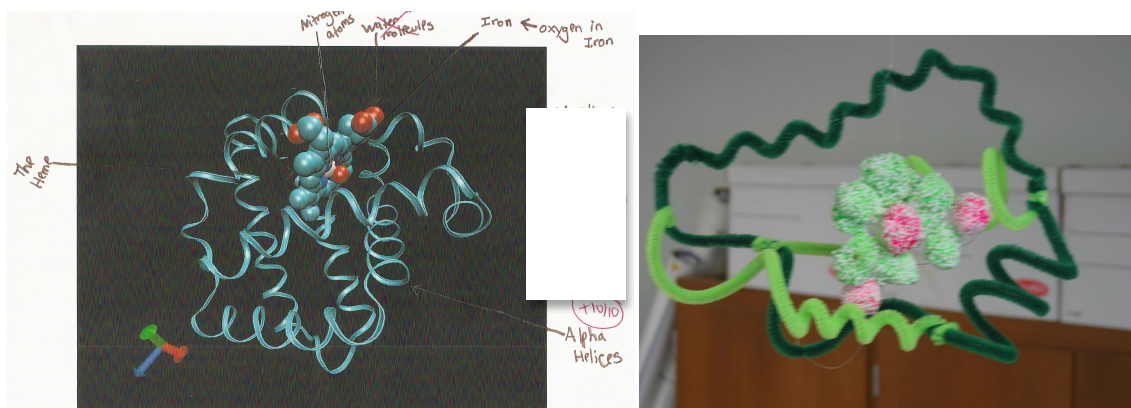
### B. Luciferin



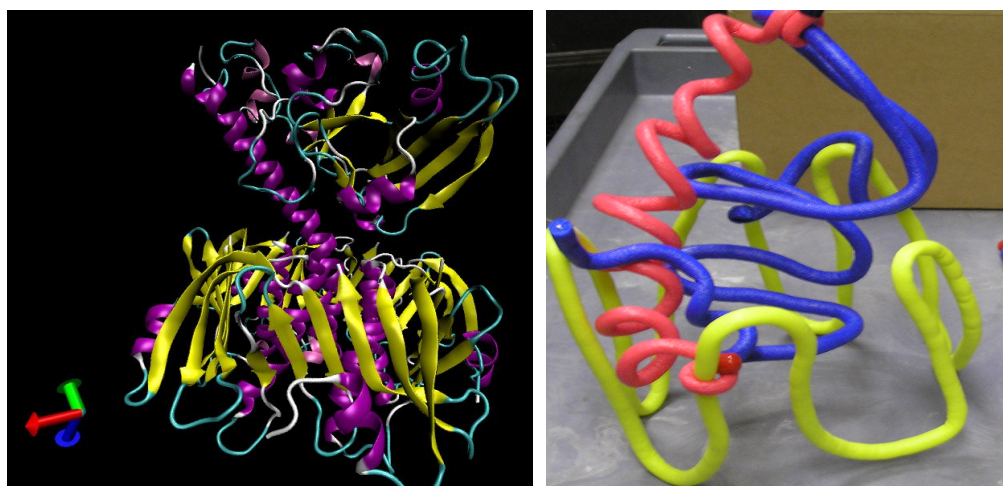
### C. ATP Synthase



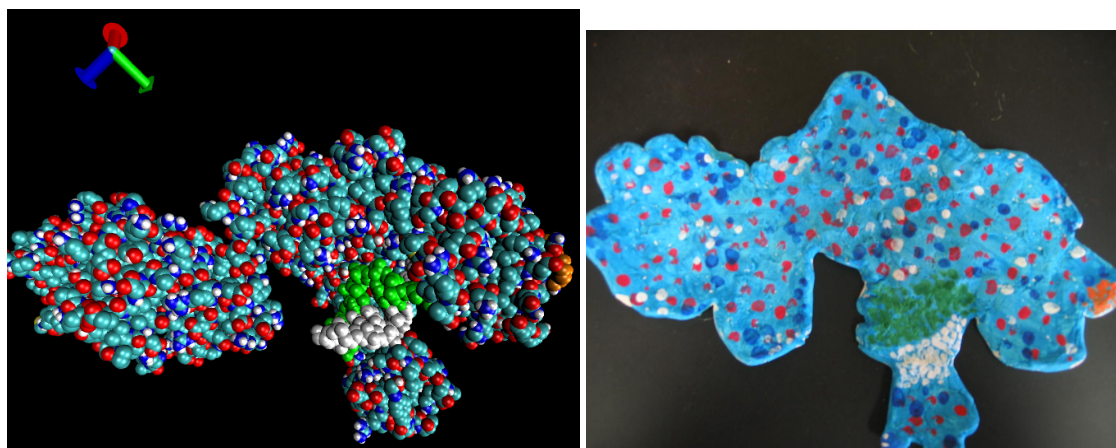
**Figure 5.12** DHS Students' VMD and Physical Models.



**A. Myoglobin**

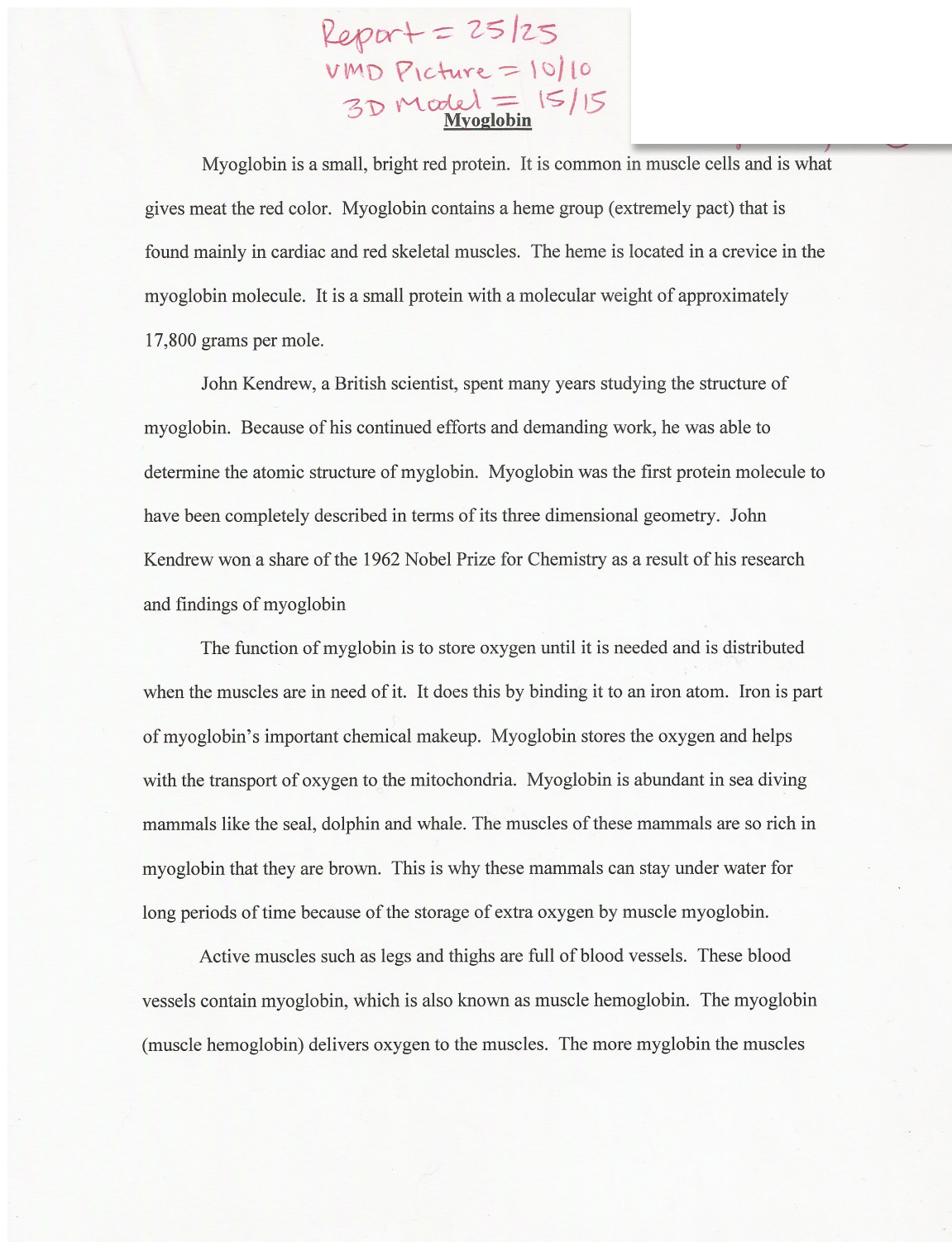


**B. Cholera Toxin**



**C. DNA Polymerase**

**Figure 5.13** DHS Student Report (continued next page).





**Figure 5.13, continued** DHS Student Report.

have, the darker the muscle. This explains why we have dark and white meat in turkeys. The breast (which is white) is rarely used and is well rested, which is why it is the color it is.

Myoglobin is a single chain globular protein containing 153 amino acids. It contains an oxygen binding heme (porphyrin ring with an iron center), around which the other apoprotein folds. The hydrogen bonds of side groups hold together the tertiary structure. It has an alpha helix. There is a proximal histidine group attached to the iron center, and a distal histidine group on the opposite face, that is not bonded to the iron. Unlike hemoglobin, this protein does not have cooperative binding of oxygen.

The structure contains one protein chain, a heme group and a sulfateior. The protein chain is composed of spring shaped alpha helices that are linked by short loops and surrounds the heme group (flat). The heme group that is surrounded by a protein chain has the iron center which usually has oxygen in it. The iron atom is surrounded by four blue nitrogen atoms.

My model represents the myoglobin structure. The dark green represents the alpha helices. The inner portion contains the heme (green) surrounding the iron center (white). The light green is hydrogen bonds holding the alpha helices together. The red represents the water molecules with the oxygen being in the center.

excellent!

~~to be~~ oxygen atoms that are part of the heme

Hydrophobic acids are those with side chains that do not like to live in a water environment. Usually you will find these amino acids buried in the hydrophobic core of the protein. Hydrophobic amino acids cannot form hydrogen bonds with water because

### **5.3.6 Grades for 2006-2007 Cohort**

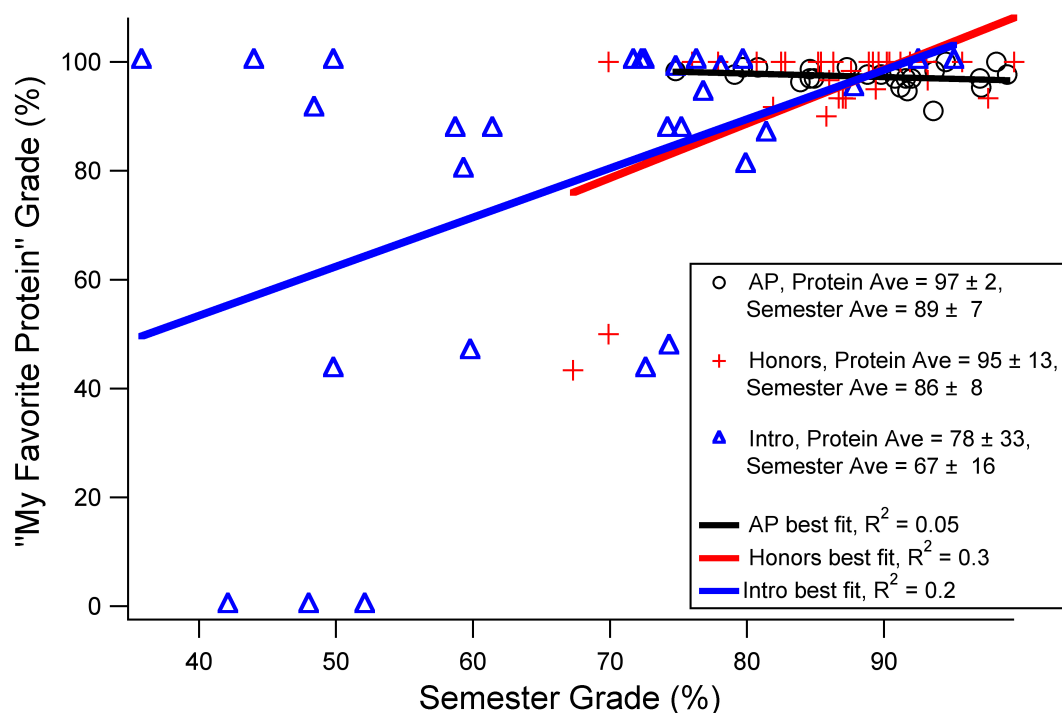
Figure 5.14 shows there is no correlation between individual student grades for the “My Favorite Protein” lesson and those student’s overall semester grades for the 2006-2007 DHS cohort. Though no student with a passing semester grade ( $>75\%$ ) received a failing grade on the “My Favorite Protein” assignment, 13 failing students received greater than 80% of points for this assignment. This includes the student with the lowest grade in all three classes who earned 100% of the “My Favorite Protein” assignment points. This result corroborates video and journal evidence that failing students appeared more engaged than usual. Overall, the introductory class grades show a much larger spread for both “My Favorite Protein” and semester grades than either honors or AP. The lowest “My Favorite Protein” grade for an AP student was 92% and only 2 Honors students received a grade lower than 90%, while half of introductory biology students received grades below 90%, not including the 3 students who did not complete the assignment (and received no credit). Similarly, all AP students had semester grades above 80% as did 5 out of 6 honors students, while less than 1 in 5 introductory biology students had semester grades that high. Overall, nearly all students performed slightly better than their semester average on the “My Favorite Protein” assignment.

### **5.3.7 Grades for 2007-2008 Cohort**

Semester grades were unavailable for this group; however, figure 5.15 shows the “My Favorite Protein” grade distribution for these students, along with grades for non-AP students from the 2006-2007 cohort. More than  $1/3^{\text{rd}}$  of students in the 2007-2008 cohort

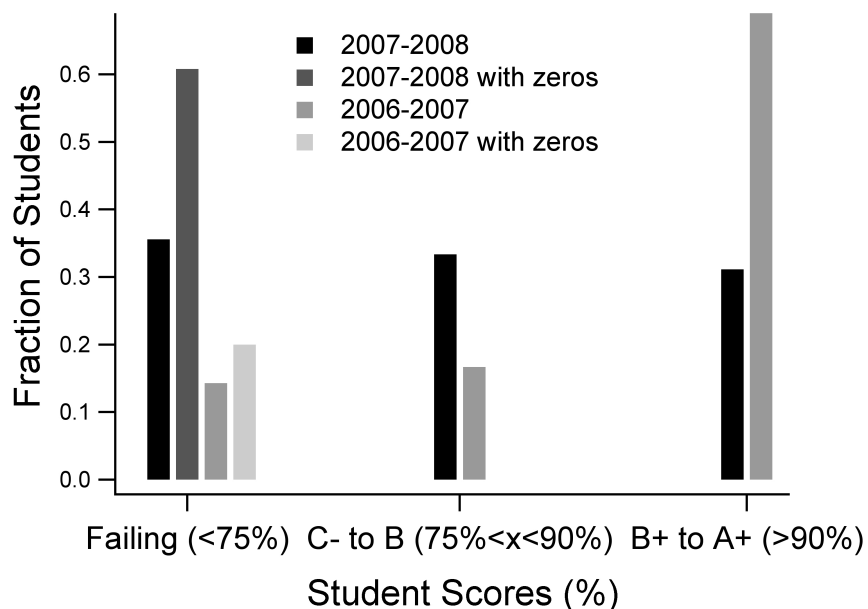
failed to turn in a model or report for the “My Favorite Protein” assignment, while only 5% of 2006-2007 students did so. Furthermore, of the students who did turn in reports, 1/3 of the 2007-2008 cohort received failing grades, while only 15% of the 2006-2007 cohort had grades that low. The primary reason for students receiving failing grades was plagiarism, which was checked via Google searches for the 2007-2008 cohort, but not the 2006-2007 cohort. This change in grading methods can account for the score discrepancy among completed assignments, but not the more than 6-fold increase in students who failed to turn anything in.

**Figure 5.14.** Semester and “My Favorite Protein” grades for DHS 2006-2007 cohort.





**Figure 5.15** “My Favorite Protein” grades for DHS 2007-2008 and 2006-2007 cohorts, excluding AP students.



## 5.4 Discussion

We have shown that CMVs employing a powerful research tool like VMD can have a positive impact on student learning and engagement on the high school biology laboratory. Though these lessons did not improve students’ understanding of molecular size and scale or the composition of matter, many students were able to synthesize lesson products that display a sophisticated understanding of proteins and the link between structure and function. We saw excellent models and reports from students at all levels, from introductory to AP and many students who ended up failing the course, nevertheless, remained engaged and produced excellent work for the “My Favorite Protein” assignment.

Overall, students enjoyed the “My Favorite Protein” activity and working with VMD, but had difficulty making sense of their protein’s structure, indicating that more

preparation is needed in this area. A simple way to address this problem is to reduce the number of possible protein choices or have the entire class work on the same protein, say hemoglobin or RNA polymerase whose structure - function relationships are much more salient and relevant to the standard curriculum.

Other than teacher journals, we did not assess the other CMV lessons and their educational impact remains unclear. It would also be useful to compare these results with a control classroom (data collected for 2007-2008, but not analyzed) and to compare DHS results to those for NVHS (data collected for both years, but not analyzed). A cursory examination of NVHS surveys reveals much more detailed responses, especially for the 2007-2008 cohort. This data may produce more useful information since students spent 50% more time completing them and were more diligent in their effort overall.

It is preferable to have NVHS survey data to compare with DHS cohorts; however, it is already clear that the academic challenges facing DHS students and teachers are much greater. This makes DHS a better environment for assessing the difficulties associated with incorporating CMVs. Lessons that can be implemented at DHS despite limitations with technology and have a positive impact despite the learning challenges presented by many unprepared, absent and unengaged students, then they are likely to be useful in the majority of biology classrooms nationwide.

## 5.5 References

1. Cook, M. P. (2006). Visual Representations in Science Education: The Influence of Prior Knowledge and Cognitive Load Theory on Instructional Design Principles. *Science Education* **90**, 1073-1091.
2. Cook, M. P., Wiebe, E. N. & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education* **92**, 848-867.
3. Atkins, P., Baumgartner, E., Chapman, O. L., Coppola, B. P., Dede, C., Garik, P., Gelder, J. I., Gotwals, B., Jakobsson, E., Johnson, A., Jones, L., Jordan, K. D., Jose, T., Kelly, R., Koel, B. E., Kosslyn, S., Kozma, R., Lewis, N. S., Linn, M., Madura, J. D., Mahaffy, P., Martz, E., Ostlund, N. S., Rowe, B., Sabelli, N., Schulten, K., Shusterman, A. J., Shusterman, G., Stillings, N. A., Tinker, R., Tversky, B., Uttal, D. H., Vining, W., Waddington, D. J., Williamson, V. M. & Wintersteen, J. (2001). *Molecular Visualization in Science Education Workshop*, Arlington, VA.
4. Jones, L., Jordan, K. D. & Stillings, N. A. (2005). Molecular visualization in chemistry education: the role of multidisciplinary collaboration. *Chemistry Education Research and Practice* **6**, 136-149.
5. Baker, S. & Talley, R. (1974). Visualization skills as a component of aptitude for chemistry: A construct validation study. *Journal of Research in Science Teaching* **11**, 95-97.

6. Wu, H.-K., Krajcik, J. S. & Soloway, E. (2001). Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching* **38**, 821-842.
7. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33-48.
8. Mathews, J. (2008). The Top of the Class 2008: The complete list of the 1,300 top U.S. high schools, In *Newsweek*, online.
9. Wefer, S. H. (2003). Name That Gene, An Authentic Classroom Activity Incorporating Bioinformatics. *The American Biology Teacher* **65**, 610-613.

## Chapter 6

### Building an Effective K-12 Outreach Program

*“Scientists and engineers working in partnerships with local teachers represent an essential new force that will be required for effective science education reform. But to be effective, we scientists must first be willing to be educated about the opportunities and problems in our schools. This means that we must approach this problem with a humility that reflects how little most of us really understand about how children learn, as well as our respect for the tremendous energy, devotion, and skill required to be a successful K-12 teacher in today’s schools.”*

— Bruce Alberts, Professor of Biochemistry and Biophysics, University of California, San Francisco and Editor in Chief, *Science Magazine*

#### 6.1 Introduction

Over the past ten years, there has been increasing attention to and funding for scientist-teacher partnerships as an approach to science education reform. In 1999, the National Science Foundation created a new graduate fellowship, the NSF GK-12 Program, to pair science doctoral students with K-12 teachers in an effort to improve communication and teaching skills among graduate students, while enriching science and math instruction for their K-12 partners. Since its inception, the GK-12 program has funded more than 200 projects at universities across the US, taking scientist participation in K-12 classrooms to unprecedented levels. The University of Illinois Urbana-

Champaign GK-12 Program ended this year after ten years of successful partnerships with K-12 schools, leaving behind a wealth of experience and materials as well as cultivated relationships with motivated local teachers. In order to preserve the work of the UIUC GK-12 Program, we created an outreach program at the Center for Biophysics and Computational Biology here at the University of Illinois. We describe our program here as a model for how faculty and graduate students can create university-K-12 outreach programs. We also review the relevant literature and current thinking about the essential features required for such programs to be successful.

## **6.2 UIUC Biophysics K-12 Outreach Program Overview**

Our outreach program pairs biophysics graduate students with local K-12 teachers in order to develop lessons which teach middle and high school students about the exciting research happening here at the University of Illinois, Urbana-Champaign. The teachers we work with come from schools in Central Illinois and suburban Chicago where they teach physics, biology or chemistry to students from 7th to 12th grade.

The graduate students are PhD candidates doing research on a vast array of topics in biophysics, from single molecule fluorescence microscopy to molecular dynamics simulation of membrane proteins. These students come from diverse academic backgrounds, including physics, biology and computer science and are eager to give back to the community while improving their ability to teach and communicate their research.

Graduate students are eligible to receive tutorial credit for developing one or more lessons in collaboration with a teacher and for publishing lesson materials on the program

website. Research faculty act as mentors and advisors to their graduate students participating in the program. In some cases, faculty may even visit the classroom or assist with lesson development and implementation.

The entire process of choosing a teacher, developing a lesson, teaching, and preparing lesson materials for online publication requires at least one month, but typically two to three months for each lesson.

To provide maximum continuity with the course, lessons are developed in collaboration with teachers and are closely tied to state learning standards and the course curriculum. Lessons are primarily hands-on, with the goal of making learning science more like doing scientific research. This means that more questions are generated than answers and activities are not canned, but rather authentic explorations that provide opportunities for both discovery and failure.

## **6.3 Essential Elements of Effective University Science Outreach**

### **Programs**

In recent years, as more and more universities become involved in K-12 science outreach, policy organizations and the science education community have attempted to determine the impact of these projects on K-12 student learning. There has been a great deal of literature on the subjects <sup>1,2,3,4</sup> as well as the development of a National Academy of Sciences/National Research Council initiative<sup>5</sup> to specifically address the needs of scientists and engineers involved in K-12 education. We review the common elements of programs that demonstrated positive learning outcomes and provide practical examples of

how to institute these elements using our biophysics outreach program as an example.

### **Meet an Existing Need.**

Successful outreach programs were consistent in their focus on meeting an existing need in the community. Ways to determine existing needs are to first read the literature and determine the significant issues in science education reform. Currently, some of the most pressing needs that university outreach programs can help address are:

1. The need to improve consistency in education across the K-16+ continuum.
2. The need for resources and best practices to support more inquiry and problem-based learning in K-12 science classrooms.
3. The need for continuing education opportunities for teachers in fields like biology, where the subject matter changes rapidly.
4. The need to incorporate the latest findings of the learning sciences into everyday teaching practices.
5. The need for diverse scientist role models, especially for disadvantaged or under-represented students.
6. The need to connect K-12 teaching to the excitement (and frustration) of discovery research.

Once program developers are educated about current issues in the field, they can meet with local teachers and community leaders to see what is already being done and determine the best ways they can help.

In developing our outreach program we benefited from feedback from the UIUC



GK-12 Program teachers. When asked what they liked about the program and what they wanted in future collaborations, teachers indicated that they wanted more exposure to current scientific research for their students and professional development opportunities that allow them to remain current in their fields. The benefits of connecting science education to research have been well-documented<sup>6</sup> and the need is still unmet; there are no programs which bridge the gap between UIUC biophysics research and K-12 teaching by direct involvement of scientists in the development and teaching of research-based lessons in local classrooms.

### **Focused Outreach.**

Once the need in local K-12 schools has been established, along with the ways in which teachers and scientists can work together to address them, program developers will need to draft a set of programmatic goals. The merit of these goals should be supported by education research, national and state learning standards, and policy reports like the AAAS Project 2061. Furthermore, each goal should be directly tied to a measureable outcome. All program activities should be justified according to how they help achieve program goals and all participants—teachers, scientists and even K-12 students—should be in agreement with and willing to support these goals. Table 6.1 shows the goals and predicted benefits of our outreach program for K-12 schools and the University of Illinois.

**Table 6.1** Program goals for the UIUC Biophysics K-12 Outreach Program (continued next page)

<b>Program Goal</b>	<b>Benefits to K-12 Schools</b>	<b>Benefits to Universities</b>
Expose local K-12 students and teachers to biophysics research occurring at the University of Illinois.	Increase enthusiasm for science, exposure to scientific careers and knowledge of emerging research.	Recruitment and training of future university students and researchers. Service opportunities for faculty.
Develop, implement and disseminate inquiry-based lessons that tie current research to standard curricula.	In addition to the above, this will improve students' understanding of the nature of science and their ability to use scientific inquiry to solve novel problems.	Dissemination of university research accomplishments.  Low cost pedagogy training for science graduate students and faculty.
Provide scientist role models for students in the classroom.	Dispel students' mistrust of science and stereotypes about who can become a scientist.	Improve the image of the university in the local community.  Help graduate students and faculty feel more connected to the community.
Improve university scientists' (graduate students and faculty) understanding of issues and challenges in K-12 science education.	More support from Universities and scientific/professional organizations for K-12 schools and science education reform efforts.	Improved understanding by scientists of how to participate in K-12 outreach. Improved understanding of the type of reforms to support at the policy level.

**Table 6.1, continued** Program goals for the UIUC Biophysics K-12 Outreach Program

<b>Program Goal</b>	<b>Benefits to K-12 Schools</b>	<b>Benefits to Universities</b>
Improve science communication and teaching skills for graduate researchers	Professional development for teachers as pedagogy experts.	Low cost pedagogy training for science graduate students and faculty.

**Well Supported.**

The importance of institutional support cannot be overstated. Principals, department heads, university faculty and administration are all central to ensuring the success of an outreach program. For a program with clear goals and evidence of the impact, gaining this support will be much easier. K-12 and university administration are similar in that there is strict oversight of time and expenditures. Teachers will feel much better about participating in an outreach program if their principals are on board with the project. Similarly, graduate students will feel more motivated to participate in outreach activities if they have support from their advisor and department head. Since outreach is time-consuming and voluntary, the projects are typically self-selecting for only the teachers and scientists who are most committed to professional development in science education. This human capital is essential, but insufficient for effective collaboration without moral and financial support from institutions. The easiest way to obtain moral support from administrators is to tie program goals to projected benefits for the institutions involved. The stated benefits should be measureable and closely tied to the institution's mission and responsibilities.

Financial support in the form of basic resources such as paper, colored pencils/pens, and more expensive laboratory supplies is also essential. Much of these materials can be borrowed from other teaching or outreach programs already at the University or in the local community. New resources can also be purchased together with an existing program for the mutual benefit of both. Student organizations can also help by competing for university funds set aside for them. University of Illinois Student Organizations Resource Fee Committee awards up to \$20,000 per year to student organizations. This type of funding should not be overlooked.

Finally, offering incentives to volunteers is highly advisable. As mentioned before, the type of teachers and scientists who will volunteer their time for outreach are already very motivated, but incentives such as course credit for graduate students and professional development credits for teachers will go a long way.

Our outreach program provides tutorial course credit for any graduate student who develops, teaches and writes up a brief report for at least one lesson in collaboration with a local teacher. Since these reports are already in the form of a “how to” paper common in some education journals, they can be compiled with other outreach lessons or submitted individually for publication. This supports outreach projects by providing students, faculty, teachers and departments with documentation of their efforts and its impact.

In the near future, we hope to provide on-site training and research opportunities for teachers that qualify for Illinois Regional Offices of Education Continuing Professional Development Units (CPDUs). Offering CPDUs is an excellent way to recruit teachers since they are directly tied to salaries.

### **Consistent & Long Term.**

Because typical university outreach programs rely almost entirely on volunteer efforts, these programs can take several years before having a significant impact. Furthermore, relationships with K-12 schools and teachers take time to build and develop into effective collaborations. For these reasons, it is crucial that an outreach program have a long-term commitment. At universities this will not be easy since the undergraduate and graduate students who typically do most of the work will eventually graduate, taking their knowledge and experience with them. K-12 institutions are also not static. It is very common, for example, for teachers to be re-assigned to new teaching duties each school year. This happens even with science teachers. A middle school science teacher can be re-assigned to teach English or History when there is a shortage of teachers in that area. A high school biology teacher can be reassigned to lower-level health science classes and physics teachers can be assigned to all math classes. In addition, teachers may move away or switch schools and, of course, K-12 students will almost never be the same two years in a row.

As such, program infrastructure is essential for maintaining continuity within an outreach program involving non-static volunteers and participants. The infrastructure can include any resources that keep the program coherent, but the most important for this type of outreach include:

- A consistent source of funds
- A program coordinator. This can be a graduate student, faculty member or departmental secretary. This will work best if some amount of funds or course

credit can be set aside for compensation in order to free up time for outreach work.

- Physical space for meetings or outreach events.
- Lab equipment for classroom lesson and demonstrations.
- Office materials and resources such as paper, printers and photocopiers.
- A program website to advertise and organize the project, to facilitate dissemination and training of new teacher and scientist volunteers, and to provide documentation of work done for funding agencies and university or K-12 administrators.
- Documentation of work. In addition to the website described above, this can include training manuals, lesson plans, resource lists (where to go to on campus to obtain certain resources, names and contact information for past volunteers, teachers and other similar outreach programs), evaluation plans and reports, and lists of equipment and materials belonging to the program. See Appendix B for our program's documentation so far.

### **Collaborative.**

The most effective teacher-scientist relationships are those that are true partnerships, where the scientists and teachers approach each other as colleagues, each with significant, complimentary expertise to bring to the project and each with a significant amount to learn. Specifically, scientists should not think of themselves only as experts who will disseminate knowledge to the teacher and students. On the contrary, many K-12 teachers are broadly knowledgeable about science and will likely already

have ideas for the best way the scientist can contribute to their classroom. The types of teachers who are drawn to outreach efforts are very motivated and are always looking for professional development opportunities. A good outreach program will clearly market itself to teachers, telling them not just how their students will benefit, but how they will benefit professionally from participation. Similarly, these types of partnerships are tremendous opportunities for scientists to improve their communication and teaching skills and learn more about the challenges in K-12 education. After all, university freshman, whom many of these scientists are charged with educating, have just emerged from a high-school environment that most scientists can't remember and know very little about. In fact, most science faculty have no training in pedagogy and little actual teaching experience. Furthermore, the types of teaching experiences many scientists' students are involved in are typically entirely didactic and adhere little to best teaching practices. Recent assessments of former graduate students in NSF GK-12 Programs reveal that these students outperform their peers in later positions as tenure track faculty and lecturers when it comes to teaching undergraduates. In light of these pedagogical benefits and the increasing importance of K-12 outreach to funding agencies, it is clear that scientists have much to gain personally from these efforts.

**Include an Evaluation Plan.**

As stated previously, evaluation is an integral part of any outreach program and is required for justification of a program's impact to institutions and funding agencies. As described above, each program goal should be tied to an expected educational benefit that

can be measured. Programs should then layout a detailed plan for how they intend to measure each impact.

Table 6.2 details the evaluation plan for the UIUC Biophysics K-12 Outreach Program. This plan can be accomplished with just two surveys and a quiz for K-12 students, along with feedback from teachers and faculty advisors. This feedback can be in the form of a simple e-mail to the program coordinator. Teacher feedback will express the benefits, limitations, and costs of the lesson from their perception, along with a comparison to existing lessons that teach similar topics. Faculty advisor e-mails should address how participation in the program has benefited them and their graduate students.

The assessment quiz for students, given before and after their exposure to scientists in the classroom, should contain three types of questions. There should be at least one question that addresses student attitudes about science and scientists (i.e. are they interested in becoming a researcher, engineer, doctor, etc. and how likely are they to believe spurious claims over scientific evidence?). Another set of questions should address *knowledge*, and at least one open-ended question to assess *understanding*. These terms are used according to the definition set forth by the National Research Council (NRC, 1996, p. 23):

*Understanding* science requires that an individual integrate a complex structure of many types of knowledge, including the ideas of science, relationships between ideas, reasons for these relationships, ways to use the ideas to explain and predict other natural phenomena, and ways to apply them to many events. Understanding encompasses the ability to use knowledge, and it entails the ability to distinguish between what is and what is not a scientific idea.



This quiz will need to be custom designed to assess basic knowledge and understanding related to the research and lesson taught.

Dissemination of lesson materials created as part of this outreach project is assessed using the simple web-form shown in figure 6.1. Using this instrument, we can clearly estimate how our materials are being used and in what contexts.

Another survey is used to assess graduate students' knowledge of issues in K-12 education and what they hope to gain from volunteering. Graduate students then complete the survey again after completion of the lesson and they reflect on what they feel they have gained and how the experience has impacted their understanding and ability to communicate their research.

**Table 6.2.** Evaluation plan for the UIUC Biophysics K-12 Outreach Program (continued next page)

Program Goal	Ways to measure impact
Expose local K-12 students and teachers to biophysics research occurring at the University of Illinois.	<p>Survey questions that assess student attitudes about science and scientists.</p> <p>Teacher feedback detailing their perceptions of the benefits and limitations of lessons.</p>
Develop, implement and disseminate lessons that tie current research to standard curricula.	<p>Brief web survey required for lesson material downloads.</p> <p>Quiz problems that assess both <i>understanding</i> and <i>knowledge</i>.</p>

**Table 6.2, continued** Evaluation plan for the UIUC Biophysics K-12 Outreach Program.

Program Goal	Ways to measure impact
Provide scientist role models for students in the classroom.	<p>Surveys questions that assess student attitudes about science and scientists.</p> <p>Survey questions that assess graduate students' knowledge of issues in K-12 education and what they hope to gain from volunteering.</p>
Improve university scientists' (graduate students and faculty) understanding of issues and challenges in K-12 science education.	<p>Research advisor feedback on the impact of the outreach experience for them and their student(s).</p> <p>Survey questions that assess graduate students' knowledge of issues in K-12 education and what they hope to gain from volunteering.</p>
Improve science communication and teaching skills for graduate researchers	<p>Teacher feedback detailing their perceptions of the benefits and limitations of the partnership.</p> <p>Research advisor feedback on the impact of the outreach experience for them and their student(s).</p> <p>Survey questions that assess graduate students' knowledge of issues in K-12 education and what they have gained from volunteering.</p>

**Figure 6.1** Web survey which tracks dissemination of program lessons

1. Have you downloaded this activity before?

☐ Yes

☐ No

2. What is your role:

☐ Volunteer for UIUC Biophysics Outreach?

☐ Volunteer for another outreach program?

Please specify\_\_\_\_\_

☐ Teacher?

What grade & subject?\_\_\_\_\_

3. What is the name of your school?\_\_\_\_\_

4. What is your city and state?\_\_\_\_\_

## 6.4 Tips for Scientists Working in K-12 Classrooms

In general, scientists who want to improve their ability to communicate their work or gain teaching skills, and who are passionate about what they do, will be great candidates for K-12 outreach. Since most students and members of the public have little to no interactions with scientists, the simple presence of a scientist in the classroom goes a long way to both dispel the inherent mistrust of science felt by much of the public and challenge stereotypes about who can become a scientist.

A scientist's main job in the classroom is not to be an expert, but to create learning activities that allow students to experience what scientists love most about

science, namely the thrill of discovery. Students will also benefit from learning about a scientist's inquiry process, his/her use of critical reasoning skills in scientific exploration, and how he/she asks the fundamental questions: *How?* *Why?* and *What if?* Figure 6.2 shows general advice for scientists working in K-12 classrooms, from the National Academies of Science Resources RISE program.

**Figure 6.2** National Academies of Science/National Research Council advice for scientists working in K-12 classrooms.

**Do:**

- observe in classrooms first
- work on breaking down the scientific language you use with colleagues
- work closely with teachers to learn pedagogy, to understand the diversity of students, to get general "reality checks"
- engage students in activities
- treat content as a way to engage students in critical thinking
- strive to become involved on a sustained basis

**Don't:**

- lecture
- take on the role of expert
- be didactic
- assume the classroom has abundant resources or equipment
- expect to be of as much help or influence--at first--as you may have hoped to be; this will develop but gradually

## **Working with Teachers.**

As most teachers have very limited time to work on outreach projects during the school year, scientists must approach meetings and deadlines formally. Rather than face-to-face brainstorming sessions or frequent e-mailing, communications with teachers should be focused, well-prepared and as brief as possible. Teachers will still be required to put in significant effort, but they will likely prefer to do this work on their own time (i.e. at home on the weekends) since school days are usually quite full.

It is important to keep in mind that classroom visits should also be focused and not overly ambitious. There are usually no more than 45 minutes of learning time in a class period and even less for labs, due to setup and cleanup. Therefore, lessons should be developed with these time constraints in mind.

Since any outreach program between scientists and K-12 teachers constitutes a new form of collaboration, there may be some difficulties. It is important that scientists respect teachers as fellow professionals and understand that while teachers may not have a deep understanding of some scientific principles, they have great expertise in terms of tailoring difficult concepts to specific age groups and skill levels. In fact, it is important to discuss with teachers which concepts their students know already and which they would be able to master. Teachers will also have substantial expertise presenting material in ways that accommodate a variety of learning styles.

While an inquiry-based approach is one of the assets that scientists can bring to the K-12 classroom, all K-12 teachers may not be comfortable with this as a learning style. Much of the experimentation that occurs at the K-12 level is meant to reaffirm what is already known; therefore the idea of an activity that doesn't produce clear results

or which leads to the conclusion that more research is necessary, may be hard for some teachers to accommodate. However, if, in the course of the collaboration, both students and teachers are able to feel more comfortable saying “I don’t know” and as a result seek to discover the answer, then this in itself is an achievement.

## **6.5 Conclusions**

In a time when the United States is facing a crisis in science education at the K-12 level, especially in terms of the disparity of educational quality between “have” and “have not” schools, universities are in a unique position to create the type of meaningful outreach programs that will act as recruiting tools for the next generation of scientists. If employed successfully, with adequate institutional support and enthusiasm by teachers, scientists, and administrators, outreach programs can show students that scientists are a diverse group of regular people engaged in exciting scientific discovery. Scientists themselves should come away from these collaborations with increased ability to teach to a range of students, communicate their research, and experience first-hand the challenges of teaching scientific principles and inquiry at the K-12 level.

## **6.6 References**

1. Tanner, K. D., Chatman, L. and Allen, D.E. . (2003) Science teaching and learning across the school-university divide—cultivating conversations through scientist-teacher partnerships, *Cell Biol. Educ.* 2, 195-201.

2. Caton, E., Brewer, C., and Brown, F. (2000) Building teacher-scientist partnerships: Teaching about energy through inquiry, *School Science and Mathematics* 100, 7-15.
3. Tanner, K. D., and Allen, D.E. (2004) Moving from outreach to partnership: Striving for articulation and reform across the K-20+ science education continuum, *Cell Biol. Educ.* 4, 69-74.
4. Tomanek, D. (2005) Building successful partnerships between K-12 and universities, *Cell Biol. Educ* 4, 28-30.
5. NRC. (2000) *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*, (Center for Science, M., and Engineering Education. National Research Council, Ed.), National Academy Press, Washington, D.C.
6. Tanner, K. D. (2000) Evaluation of Scientist-Teacher Partnerships: Benefits to Scientist Participants, in *National Association for Research in Science Teaching Conference*. New Orleans, LA.

## **Appendix A**

### **Laboratory Protocols**

#### **A.1 SUV Prep Protocol**

Procedure:

- 1) Use a volumetric, glass pipet to deliver the desired amount\* of lipids into a round bottomed flask (RBF). Make sure to rinse both the flask and the pipet with pure chloroform prior to dispensing lipids. Important: Do not allow chloroform (including lipids in chloroform) to come into contact with any plastics.
- 2) Dry under a nitrogen stream while rotating the flask to leave a uniform lipid film around the flask interior. Remove all residual chloroform by lyophilizing (quick freeze in liquid nitrogen and then apply vacuum to at least 500 milliTorr for at least 2 hours).
- 3) Add enough room temperature buffer to the flask to achieve the desired lipid concentration (For tip sonication, at least 4ml of resuspension buffer is required since the most narrow tube which will still accommodate the tip thickness requires 4mL of liquid to reach the necessary 2-inch height for successful sonication). Vortex suspension on low power and swirl for several minutes to resuspend all lipid from the walls of the flask. During mixing it is important not to generate bubbles. Hydration time may differ slightly among lipid species and structure; however, a hydration time of 1 hour with vigorous shaking, mixing, or stirring is sufficient.
- 4) Transfer the vesicle suspension into polypropylene or polystyrene tubes in aliquots of 10ml or less. Ensure tubes are properly sealed, but not air tight, before freezing the vesicles by plunging them into liquid nitrogen. Incubate vesicles in a 45-50°C water bath to thaw. Repeat this cycle at least 4 times.
- 5) Pool all aliquots into one large polystyrene or polypropylene container. For tip sonication, the liquid must be at least 2 inches high in the tube and the tube must be at least 1 cm thick.



- 6) For sonication, use the Sonics & Materials, Inc. Vibra Cell with horn #? (Morril Hall 5<sup>th</sup> floor). Sonication is pulsed with a 60% duty cycle and output control level of 5. Always make sure to rinse the sonicator tip with ethanol before dipping it into the sample. To prevent overheating, suspend the sample in a beaker of ice water and cool the tip using an argon stream if it is available. Sonication time depends on the concentration of lipid as well as the total volume of the suspension. An SUV sample with a dominant population of 30-50nm diameter vesicles can be generated from sonicating a sample composed of 3mM POPC for 15-20 minutes.
- 7) Large particles, including any titanium that may have sloughed off from the sonicator tip, must be removed by ultracentrifugation. Use the Beckman TL-100 benchtop model with the TLA100.3 rotor (5<sup>th</sup> floor Morril Hall) and the Beckman 3mL polyallomar quick-seal tubes (part number 349621). Make sure to balance each pair of tubes along with their spacers (Beckman catalog #355937) prior to spinning. To quick seal, allow the heat button (5<sup>th</sup> floor Morril hall) to warm up at least 10 minutes. Fit the seal former cap (Beckman catalog #348120) onto the tube top and position the tube in the holder underneath the heat button. Press down firmly on the button and hold until the tube neck has almost disappeared. Release the button and allow the seal to cool for 1 minute. Remove seal cap and invert tube while squeezing to ensure the seal is complete.
- 8) Spin the samples for 2hrs at 55,000 rpm (~125,000xg) and 10C. Ensure the rotor is not too cold ( $\leq 5^{\circ}\text{C}$ ) before you load the sample tubes.
- 9) Remove the upper 2/3 of centrifuged sample in the following manner: a) Place a 21 gauge needle into the side of the tube opposite the pellet, at 1/3 of the way up from the bottom (be careful not to poke and release the pellet). Attach another 21 gauge needle to a 10 or 20cc syringe and use this to puncture the top of the tube where there is a small air pocket. Use this syringe to force air into the tube and push the liquid out of the bottom syringe and into a sterile container placed there to catch it.

- 10) Pool the centrifuged supernatant and mix thoroughly before filtering using an Avanti Mini-Extruder. Pass lipid suspension ~17-21 times through 2 stacked 50nm pore polycarbonate filters. Pool all collected supernatant and store at 4C.
- 11) The size distribution of each batch of SUV should be checked using Dynamic Light Scattering and a qualitative assessment of size and lamellarity can be obtained from negative stain Transmission Electron Microscopy (see vesicle characterization protocol). If the sample is not monodisperse, further filtration can be performed.

\*It is estimated that each lipid in a fully hydrated POPC bilayer occupies 64 square angstroms (Gullingsrud et al, Biophys. J. 2004 (86) 3496). This means that 12,026 lipids are required to make a 35nm diameter vesicle and 721.6 micromoles in 20mL are required to produce 35nm vesicles at a 3uM concentration.

## **A.2 Dynamic Light Scattering**

### **Reagents & Equipment:**

1. Brookhaven BI-9000AT Autocorrelator, a 516nm laser with detector at 90 degrees and water bath controlled sample temperature. This is in room 313 MRL (Zukowski group).
2. 10x75mm Fisher disposable culture tubes (product number 14-961-25)
3. 1.0mL pipetman and tips.
4. Ethanol
5. KimWipes
6. Parafilm

Procedure:

1. Prepare 10x75mm borosilicate culture tubes by rinsing them with double DI water and then ethanol, followed by drying upside down in a test tube rack on top of sterile KimWipes.
2. Follow Zukowski group rules for instrument start-up, shutdown, and operation.
3. Prior to dispensing sample, blow out pipet tips and tubes using a bottle of compressed air.
4. Use the Brookhaven DLS software with the dust filter on and the best delay times to fit the data (usually 1msec 1<sup>st</sup> delay and 3msec last delay, but it's very dependent on particle size and concentration). Ensure the temperature is near 25°C and is accurately entered. Scan for 15 minutes.
5. Analysis: A monodisperse sample will give a mean effective diameter within 5nm for all of the fits except the linear fit, will have a polydispersity less than 0.2, and a base difference less than 0.1%. Use the analysis tools to determine if there are any large particles >300nm in the sample. Remember to display results based on number of particles, not intensity, since larger particles will scatter light more intensely than smaller particles. If the sample is not monodisperse, further filtration can be performed.
6. Saving data: Copy each of the four data windows (counts plot, correlation function plot, cumulants analysis, and run parameters) to the clipboard ("clip" at the top of the window) and paste them as bitmaps into a word document. Other analysis windows can also be included, but only for fits that approximate the correlation function reasonably well.

### **A.3 Negative Stain Transmission Electron Microscopy**

Reagents and Equipment:

1. Fresh (< 2 months since purchased) 300 mesh carbon-coated copper (#CF300-Cu, Electron Microscopy Sciences). These are glow discharged just before shipping and will begin to lose their charge after about a month. Bad discs can be detected

by a partially or completely peeled off coating on the dull side of the disc. If the coating is completely gone, the disc will appear uniformly black.

2. Tweezers with a fine point.

3. Stains:

- 2% Ammonium Molybdate pH 6-6.5

Make with pure water. Allow to stand 1 hour before bringing solution to the correct pH. Prepare fresh.

- 2% Phosphotungstic Acid pH 6.1-6.5

- Make with pure water. Allow to stand 1 hour before bringing solution to the correct pH. Prepare fresh.

- Saturated Uranyl Acetate. No pH adjustment required.

Make with pure water. 7% is saturated enough. Filter just before use.

Cover with foil and store at room temperature for no more than 10 days.

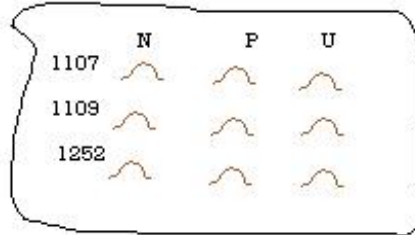
4. You will need a Center for Microanalysis of Materials user number and account (received after training completed on the Philips CM12 TEM). Analysis will take at least 4 hours.

Procedure:

1. Vesicle suspensions should be approximately 3mM lipid to ensure sufficient detection and resolution.
2. To apply your sample to the grids, follow the protocol in figure 1.

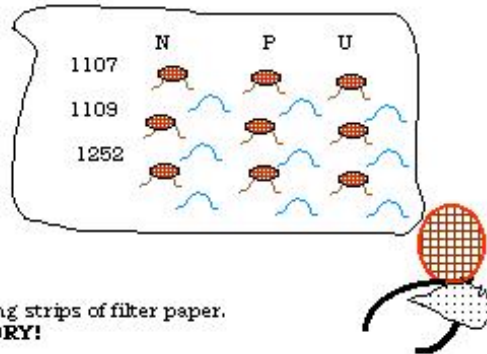
**Figure A.1** Adding sample to TEM grids (from the Center for Microscopic Imaging, UIUC).

Label areas on a parafilm sheet for each sample and each stain. Use one drop of sample for every stain to be used, ie:



Pipet 15 microliters of sample for each grid to be stained. Incubate a plastic coated grid, plastic side down facing the drop for 5-10 min, depending on the sample density.

Pipet desired stains beside appropriate floating grids.



Gently blot grids from the side, using strips of filter paper.

**DO NOT LET THE GRID DRY!**

Place the grid face down on the stain. Stain 1 min for UA, and 2 min for PTA and NH<sub>4</sub>M. **NEVER STAIN LONGER THAN 2 MIN. THE GRIDS WILL OVERSTAIN!** Clean the forceps in alcohol inbetween each sample.

\*\*\*\*\* Routine dx virus, just do the NH<sub>4</sub>M, one grid. Repeat with other stains only if there is a problem.

Pick up grid, blot from the side and place in a grid box. Cover with one chip of Drierite. Dry 20 min. Remove Drierite.

*\* Too much drierite, or leaving the drierite on the grid box too long can cause dust contamination.*

3. Follow the CMM instruction manual for Philips CM12 Alignment and operation.

## A.4 Total Phosphorous Assay

Equipment:

Hewlett-Packard 8452A Diode Array Spectrophotometer

Computer with Hewlett-Packard HP UV/VIS Chemstation Software

HELLMA 1cm path UV Quartz Sample Cell (QS 1.000)

KimWipes

10x75mm borosilicate culture tubes (cat# 14-961-25) (need 11 + 3\*(# of samples))

1LVolumetric Flask

2 x 100mL beakers

ddI water

2 x 100mL Amber glass bottles

50mL falcon tube

Tinfoil

Heating block to fit at least 15 tubes

Thermometer (capable of measuring to 300C)

Hot plate

Reagents:

Deionized water

Conc. H<sub>2</sub>SO<sub>4</sub> (Mallinckrodt), cat# 2876

Concentrated perchloric acid

Ammonium molybdate(VI) tetrahydrate (Aldrich cat. no 22,123-6)

L-Ascorbic acid (Aldrich cat. no. 25,556-4)

0.05mg/ml (1.613mM) phosphorus standard solution

Hydrogen peroxide (Fisher cat. no. H323-500)

Procedure:

- 1). Place the heating block with thermometer on the hot plate. Begin preheating block (it will take some time to obtain a stable 200°C temperature).
- 2). Prepare the solutions:
  - a. 4:1 H<sub>2</sub>SO<sub>4</sub> in perchloric acid. Make 1.2ml for the standard curve plus 0.3mL per lipid sample to be analyzed.

- b. 10% Ascorbic acid solution. Place 2 g of ascorbic acid to a 50mL falcon tube and QS with ddI water to 20mL. Mix the solution well. Store covered in tinfoil at 4°C for up to 1 month.
  - c. 2.5% Ammonium molybdate(VI) tetrahydrate solution. Place 0.5 g of ammonium molybdate(VI) tetrahydrate to a 50mL falcon tube and QS with ddI water to 20mL. Mix the solution well. Store covered in tinfoil at 4°C for up to 1 month.
  - d. Phosphorous standard. Prepare using  $\text{Na}_2\text{HPO}_4$  (142g/mol) or  $\text{KH}_2\text{PO}_4$  (136g/mol). Make a 1.613mM solution in 1L using volumetric flasks and pipets to minimize error. Adjust pH using HCL or NaOH to 7.0. Store at 4°C for up to 12 months.
- 3). Place sample (15-30 nanomoles phosphorous) into the bottom of each tube. When calculating the expected concentration of phosphorous in the sample, do not forget to include the contribution from any buffer components. Gently remove any solvent from the tubes with  $\text{N}_2$ .
- 4). Prepare the 5 standards in duplicate by placing the following quantities of phosphorus standard into the tubes: (A) 3.23 nmoles (2  $\mu\text{l}$ ), (B) 16.13 nmoles (10  $\mu\text{l}$ ), (C) 32.26 nmoles (20  $\mu\text{l}$ ), (D) 80.65 nmoles (50  $\mu\text{l}$ ), and (E) 161.29 nmoles(100  $\mu\text{l}$ ). Also include 2 blanks.
- 5). Add 100uL of the acid mixture to each sample, standard, and blank.
- 6). Cover each tube lightly with tinfoil and place them in the 200°C heating block. Bake the samples for 2-3 hours until they are colorless.
- 7). Turn the hot plate down or off (we are trying to cool to 100°C). Remove the samples from the heating block and place them at room temperature ~5 minutes before adding 1.5mL ddH<sub>2</sub>O, 200uL 2.5%  $\text{NH}_4\text{Mo}$ , and 200uL 10% Ascorbic Acid. Vortex to mix.

- 8). When a stable 100°C temperature has been reached in the heating block, place covered samples back in the block and boil for 1 hour.
- 9). Remove samples from the heating block and turn off hot plate. Allow samples to cool on the benchtop ~30minutes.
- 10). Analyze samples using a spectrophotometer as follows:
  - e. Zero the spectrophotometer using DI water.
  - f. Determine the absorbance of each blank, the five standards, and each sample at 820 nm.
  - g. Generate a calibration curve using the standards and determine the concentration of phosphorus in the samples.
- 11). References:
  - i. Sokoloff, L., Rothblat, G. H. 1974. *Proc. Soc. Exp. Biol. Med.* 146:1166-1172.
  - ii. Chen, Toribara, and Warner (1956) *Anal. Chem.* 28:1756-1758.
  - iii. Fiske & Subbarow (1925) *J. Biol. Chem.* 66:374-389.
  - iv. Bartlett, G.R., (1959) "Phosphorus Assay in Column Chromatography," *Journal of Biological Chemistry* 234:466-468.

## **A.5 Membrane Peptide Characterization: Experimental Outline**

1. Solubility Test: Resuspend lyophilized peptide in pH 7.0 phosphate buffer to a concentration of 1-20uM. Determine solubility by qualitative observation of solution clarity and UV/Vis absorption spectra of sample. Aggregated protein will appear cloudy and will show a broadened peak at 280nm as well as increased intensity in the near UV (greater than 300nm). Peptides that are aggregate at pH 7.0 will be tested for solubility at pH 5.0, 6.0, & 8.0. In the event that some



- aggregation is present at all pH values, samples will be tested for solubility in buffers of increasing ionic strength, from 50 to 500mM phosphate. Further testing will be performed at the pH closest to 7.0 and at the lowest salt concentration in which the sample remains clear and where UV/Vis spectra show very little scattering. If no such conditions can be found, the peptide will not be further characterized.
2. Solubility Limit: To assess the secondary structure of the peptide and to reveal any aggregation, CD spectra will be taken of 1uM, 5uM, 25uM, & 125uM peptide. If a change in secondary structure is observed between two concentrations, then aggregation has occurred. More measurements should be made at concentrations in the transition region (e.g. between the two concentrations where the 1<sup>st</sup> change was observed) to determine the exact limit of solubility.
  3. Membrane Binding & Structure Tests: CD spectra from 200-250nm will be taken at room temperature with and without SUV. Use the lowest protein concentration and the highest SUV concentration you can (i.e. without maxing out the HT voltage). With the 1mm path cuvette, this should be ~0.5nM SUV. If no secondary structure change is observed upon adding the SUV, then binding has probably not occurred and charged vesicles should be tried. Prepare SUV with 1%, 5%, & 25% PG in PC and repeat the above measurements. If no change is observed at any of these conditions and the peptide contains a tryptophan, then further characterization is required – proceed to step 4. If conditions are found at which a change is observed, then a melt should be taken – proceed to step 5.
  4. Membrane Binding Test (Trp-containing peptides only): Fluorescence spectra of 1-20uM protein will be taken with and without SUV. Use the lowest protein concentration and the highest SUV concentration you can – i.e. when the SUV signal comprises about 25% (this will usually be ~500nM SUV for 5uM protein). If a shift of at least 5nm is observed in the fluorescence maximum, then sufficient binding has occurred. If no shift or a negligible shift is observed, then charged vesicles should be tried. Prepare SUV with 1%, 5%, & 25% PG in PC and repeat the above measurements. If conditions are found at which a >5nm spectral shift is

observed then a melt should be taken, proceed to step 6. If no binding is observed at any of these conditions, then the peptide will not be further characterized.

5. Temperature Tunability of Membrane Binding by CD: Obtain CD spectra from 200-250nm of 4 samples: the peptide in buffer at 5°C, the peptide in buffer at 95°C, the peptide with SUV at 5°C, and the peptide with SUV at 95°C. As in step 3 you will use the lowest protein concentration and the highest SUV concentration you can. You want to determine whether there is a difference between the temperature-induced structure changes for the peptide alone in solution and the changes observed in the presence of SUV. If a change is observed, determine the melting temperature by monitoring the 222nm CD signal at 5°C intervals over the 5-95°C temperature range. Regardless of the results obtained here, go to step 4 if the protein contains a tryptophan.
6. Temperature Tunability of Membrane Binding by Fluorescence (Trp-containing peptides only): 4 samples will be measured in parallel using the multicell holder: peptide with SUV, SUV reference, peptide without SUV, & buffer reference. As in step 4, you will use the lowest protein concentration and the highest SUV concentration you can. Take a melt from 5 to 95°C, increasing the temperature in 5°C increments. Any melt that begins below 20°C requires purging the sample chamber with N<sub>2</sub> gas. SVD will be used to analyze the data. Samples which do not exhibit a binding transition over the 5-95°C range cannot be used any further.

\* Note: You will use POPC vesicles unless otherwise specified. POPG will be used whenever charge is needed. If the peptide is under 25 residues, PC lipids with shorter fatty acid chains (e.g. DMPC & DMPG) will be chosen. In the event that a different lipid system is chosen, care will be taken to avoid lipid phase transitions.

## **A.6 Gel electrophoresis (PAGE)**

1. Assemble the glass slab for gel electrophoresis and fill with ethyl alcohol to check whether it leaks. Air clean ethyl alcohol.
2. Prepare Gel.

Acrylamide (CAUTION: CAUSES CANCER!)

APS(Ammonium peroxodisulfate) is stored in 4°C, warm it up with heat block for one minute.

Afterward add TEMED, shake the tube vigorously and fill in to the plates in 1min.

Fill the solution to around 2/3 height of the glass plate.

Cover the gel solution with ethyl alcohol.

Wait for 10 minutes for the polymerization.

Pour the ethyl alcohol away and air clean it.

Prepare the stacking gel. It use less acrylamid and will polymerization slowly.

Fill the plate with the gel and insert the plastic comb.

Wait for 10 minutes for the polymerization and take the plastic comb away.

Indentations for containing of samples form.

3. 10uL standard sample stored in -20°C, mixed with 10uL violet loading buffer.  
Add 20uL sample proteins solutions and 20uL loading buffer(SDS) into a small tube and heat it up to 80°C for 2-3 minutes to denature the protein.
4. Fill the standard and protein sample into the chambers. No more than 15uL and fill it slowly.
5. Fill 15mL cathode buffer+60mL distilled water on the upper sink.
6. Fill 20mL anode buffer +180mL distilled water on the bottom sink.
7. Set the voltage around 120V and run for around 2-3 hours.
8. Took the sheet of mino- gel and rinse it with 100mL distalled water 3 times fro 5 min.  
Remove SDS and buffer salts, which interfere with the binding of dye to the protein.
9. Add SimplyBlue™ Safe Stain 20mL to cover the gel. Stains the gel for one hour at room temperature with gentle shaking. This needs to be done quickly after the electrophoresis finishes in order to prevent the band disusing.
10. Discard the stain and add 100mL distilled water. Shake it gently for one hour. Take a final picture of it.

#### Safety Notes:

- Acrylimide is extremely toxic, causing central nervous system paralysis. It can be absorbed through unbroken skin. If skin comes in contact with acrylimide solution or

powder, wash immediately with soap and a lot of water. Unpolymerized acrylimide should be polymerized with excess catalyst and disposed of with solid waste. DO NOT POUR UNPOLYMERIZED ACRYLIMIDE DOWN THE SINK.

- Amonium Persulfate should be made up fresh or used from a relatively fresh stock. It goes bad after a week or two in the refrigerator. It can be disposed of by dilution with water and pouring down the sink.

- TEMED should be stored in the refrigerator in dark glass bottles. A bottle should be good for about a year, maybe longer.

Gel Buffer. 3M Tris base and 3mM SDS solution.

Dissolve 181.7 g Tris base and 1.5 g SDS in 350 mL MQ H<sub>2</sub>O. Adjust the pH to 8.45 using concentrated HCl. Add Tris base step by step. Bring to a final volume of 500 mL with MQ H<sub>2</sub>O. Store at 4 °C.

Anode Buffer (Lower Buffer). 0.2M Tris base buffer.

Dissolve 12.11 g Tris base in 400 mL MQ H<sub>2</sub>O. Adjust to pH 8.9 with HCl. Bring to 500 mL with MQ H<sub>2</sub>O. Store at 4 °C

Cathode Buffer (Upper Buffer). Dissolve 6.055 g Tris base and 8.96 g Tricine in a total volume of 500 mL MQ H<sub>2</sub>O. Add 0.5 g SDS. Store at 4 °C.

## **A.7 Fluorescence Anisotropy using the Varian Cary Eclipse Fluorometer**

*The following instructions were obtained via email communication from Sherry Hemingson at Varian, Inc.*

General Notes:

1. To maximize your signal to noise levels and the validity of your data, whenever possible, try to keep the difference between the excitation and emission wavelengths used for measurement to at least twice the sum of the excitation and emission slits. So if you are exciting at 485 nm and viewing emission at 515 nm, with 5 nm slits for excitation and

emission (sum = 10 nm), this should be fine since there is at least 20 nm difference between excitation and emission wavelengths. It is ok to get as close as the sum of the slits, but there is more chance for scatter to contribute to the measurement, depending on the level of signal.

2. If you are inputting your own sample names into the section showing samples, you can edit the name or use the defaults, however, editing is different if you are in Advanced Reads vs. Scan. If you are in Scan, you simply click once to highlight the name and then right click and select “Change Name.” Once this is done, a field comes up that you can type the new name in and hit enter. However, in Advanced Reads, you simply click once on the sample name and start typing to enter the new name.

#### Instructions for Using the “Automated Polarization Measurements in Scan” ADL

Click on the Icon “Eclipse ADL Program Selector” (which may also be called “ADL Program Selector”). This should open up the Program Selector window. In some cases, it only opens the ADL shell as a minimized icon at the bottom of the screen and simply clicking on the icon on the toolbar opens the program.

In the Program Selector window, click once on “Automated Polarization Measurements in Scan” to highlight it and then click on OK at the bottom of the window.

Click on the Setup button on the right of the Application window that is now visible and enter the necessary parameters for your scan in the fields under the Cary and Option tabs. DO NOT click on the Polarizer button on the Accessory tab, as this is only to operate the manual polarizers. Under the Reports tab, edit the fields you wish and always check the parameters box. Click OK at bottom when finished.

Now a window titled “Automatic Polarizer Measurements” comes up. Click on “Always measure G Factor” if you want to measure this value before each sample. (not a bad idea)

Check whether you want to measure Anisotropy, Polarization, or Total Intensity. The easiest and recommended setting for the high voltage for the PMT detector is to select Auto, which will scale the signal to the appropriate high voltage for both the measurement of the G Factor and the sample scan. Alternatively, the PMT voltage can be set to Manual and enter the same value for both. Most important is to set the Voltage for sample measurements so that you are getting enough signal. The  $I_{HH}$  and  $I_{HV}$  values that are measured to calculate the G Factor are usually much higher than the  $I_{VV}$  and  $I_{VH}$  values used to measure polarization or anisotropy, so you can set the voltage for the G Factor measurement lower, if necessary. BE SURE that you have the polarizers in the automated polarizer accessory in the positions you select in this window. It is relatively simple to double check if you are not sure and the validity of your measurement depends on it! Click on OK.

To start the acquisition, DO NOT click on the Start Button at the top of the screen. Click on the Collect button on the side bar.

You will first be prompted to put in a fluorescent sample and can use your sample for this. A series of scans will be collected in order to calculate the G Factor. Make sure the values are on scale (<1000).

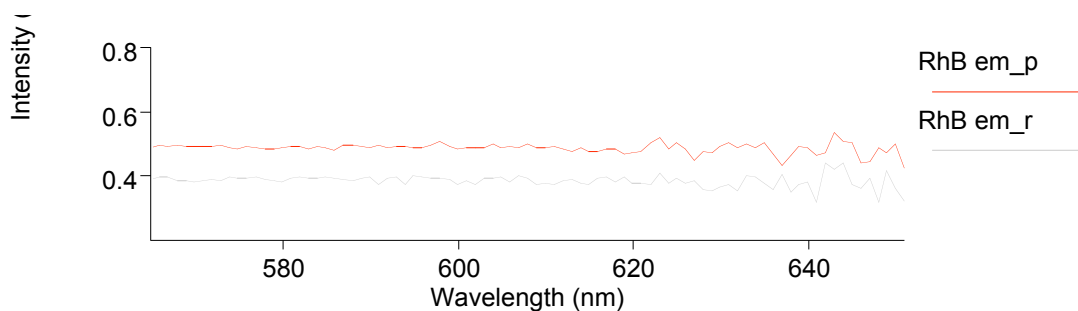
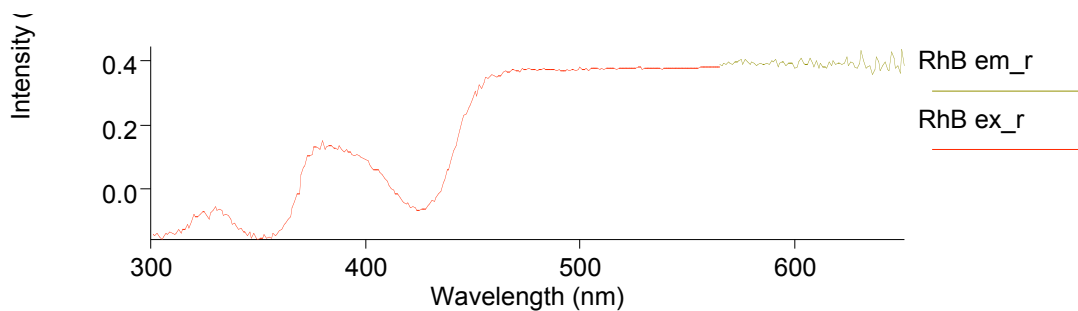
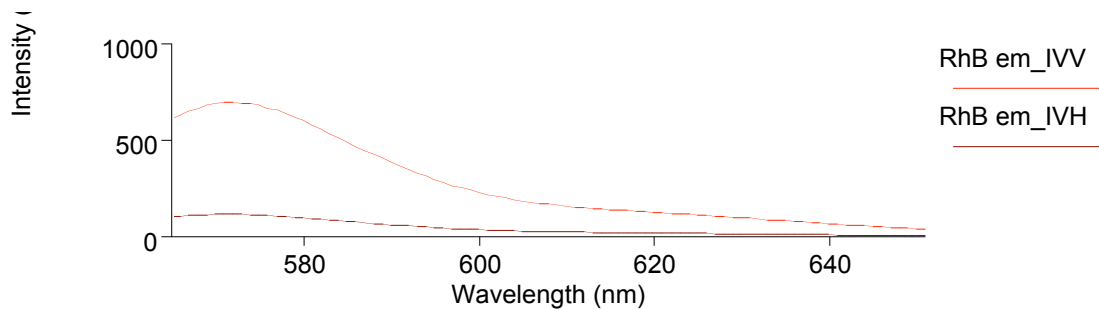
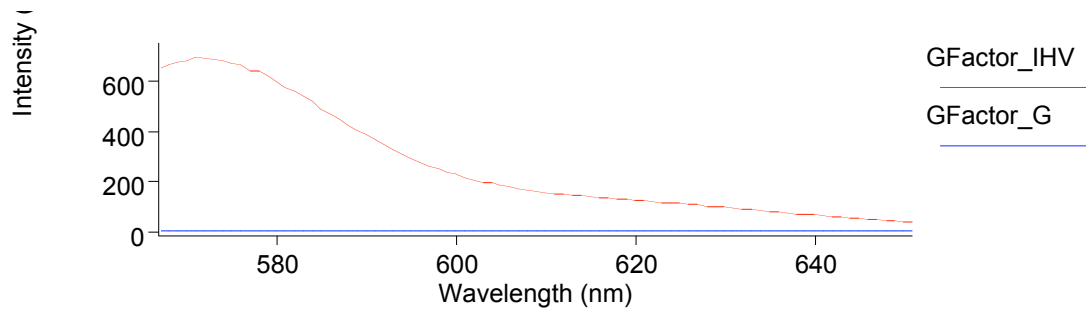
- >If you are doing an emission scan, a G Factor will be calculated for each emission wavelength.

- >But, if you are doing an excitation scan, the program will ask you to select the region over which you would like to calculate the G Factor, by clicking on the graph on the screen to mark off the range to be used for the calculation. Select a region that has the best fluorescence signal. A G factor will then be calculated.

Then you will be prompted to place your sample in the compartment and provide a sample name. Upon selecting OK, the run will start. Sample spectra and data are provided for excitation and emission scans of Rhodamine B in a PMMA block in the attached files.

Here are spectra collected for anisotropy measurements of Rhodamine B in a PMMA block using the Scan ADL. Three graphs will appear, unless you select overlay

spectra. The values for the anisotropy,  $r$ , shown in the third graph, approach that for a fluorophore of very restricted motion, around 0.38-0.39. Values for the polarization,  $p$ , have a higher upper limit of 0.5 due to the nature of the calculation, and the fourth graph shows the Rh B in PMMA sample approaching that value.



### Using the “Automated Polarization Measurements in Advanced Reads” ADL

1) Click on the Icon “Eclipse ADL Program Selector” (which may also be called “ADL Program Selector”). This should open up the Program Selector window. In some cases, it only opens the ADL shell as a minimized icon at the bottom of the screen and simply clicking on the icon on the tool bar opens the program.

2) In the Program Selector window, click once on “Automated Polarization Measurements in Advanced Reads” to highlight it and then click on OK at the bottom of the window.

3) Click on the Setup button on the right of the Application window that is now visible and enter the necessary parameters for your scan in the fields under the Cary and Option tabs. Notice that the “User Collect” box is checked under the Cary Tab, which runs the ADL from the Setup window. Although the default averaging time is 2 s, you may find that you need to increase that to improve signal-to-noise, depending on the strength of your signal. You will also need to adjust the PMT high voltage so that all the combinations of  $I_{HH}$ ,  $I_{HV}$ ,  $I_{VH}$ , and  $I_{VV}$  are on scale, and not too small in the case of the  $I_{VV}$  and  $I_{VH}$  values. This is not done automatically in the Advanced Reads ADL, so you may have to try a few runs to optimize the high voltage parameter. DO NOT click on the Polarizer button on the Accessory tab, as this is only to operate the manual polarizers. The default for number of samples is FOUR under the Samples tab. You will need to adjust this to the number of your samples; if you prefer to do statistics, you can set up replicates under this tab, also. Under the Reports tab, edit the fields you wish and always check the parameters box. Click OK at bottom when finished.

4) Now a window titled “Automatic Polarizers” comes up. Click on “Always measure G Factor” if you want to measure this value before each sample. (not a bad idea)

5) Check whether you want to measure Anisotropy or Polarization and enter the wavelength pairs for excitation and emission (up to six pairs). Also be sure to verify that the polarizers are in the positions you select on this screen. Click OK.



6) To start the acquisition, click on the Start Button at the top of the screen.

7) You will first be prompted to put in a fluorescent sample and can use your sample for this. Make sure the values are on scale (<1000). Then you will be prompted to place your sample in the compartment and provide a sample name. Upon selecting OK, the run will start.

Sample data files are also available for the examples that follow.

In the next several pages are some sample data sets for collecting anisotropy and polarization data for Rhodamine B in a PMMA block.

## Set 1: Anisotropy data collected for one sample with no replicates. Advanced Reads Report

Report Time : Thu 18 Apr 05:20:51 PM 2002  
Software Version: 1.0(83)  
Operator:

### Instrument Parameters

Instrument	Cary Eclipse
Instrument Serial Number	e100083670
Data mode	Fluorescence
User Result	execute("AutoPolarizationCollect.ADL")
Ex. Slit (nm)	5
Em. Slit (nm)	5
Ave Time (sec)	5.0000
Excitation filter	Auto
Emission filter	Auto
PMT Voltage (V)	Medium
Replicates	1
Sample averaging	OFF
Comments:	

### G-Factor

Instrument	5			
Data mode	Fluorescence			
Ex. Slit (nm)	5			
Em. slit (nm)	5			
Ave. time(s)	5.00000			
Ex. WL (nm)	Em. WL (nm)	G-Factor	Int(HV) (a.u)	Int(HH) (a.u.)
555.00	575.00	2.0804	623.943	299.917

### Analysis

Collection time 4/18/02 5:21:08 PM

## Anisotropy

Sample Name	Ex. WL (nm)	Em. WL (nm)	r	G-Factor	Int (VV)	Int (VH)
Sample 1	555.00	575.00	0.39	2.0804	126.170	20.894

The above r value corresponds well with the expected anisotropy for a fluorophore in a very rigid environment.

## Set 2: Anisotropy data collected for one sample with 3 replicates.

## Advanced Reads Report

Report Time : Thu 18 Apr 05:23:17 PM 2002  
Software Version: 1.0(83)  
Operator:

### Instrument Parameters

Instrument Cary Eclipse  
Instrument Serial Number e100083670  
Data mode Fluorescence  
User Result execute("AutoPolarizationCollect.ADL")  
Ex. Slit (nm) 5  
Em. Slit (nm) 5  
Ave Time (sec) 5.0000  
Excitation filter Auto  
Emission filter Auto  
PMT Voltage (V) Medium  
Replicates 3  
Sample averaging OFF  
Comments:

### G-Factor

Instrument	5
Data mode	Fluorescence
Ex. Slit (nm)	5
Em. slit (nm)	5
Ave. time(s)	5.00000

Ex. WL (nm)	Em. WL (nm)	G-Factor	Int(HV) (a.u)	Int(HH) (a.u.)
555.00	575.00	2.0693	624.366	301.731

### Analysis

Collection time 4/18/02 5:23:34 PM

### Anisotropy

Sample Name	Ex. WL (nm)	Em. WL (nm)	r	G-Factor	Int (VV)	Int (VH)
Sample 1	555.00	575.00	0.39	2.0693	125.833	20.917
Sample 1	555.00	575.00	0.39	2.0693	126.133	20.972
Sample 1	555.00	575.00	0.39	2.0693	125.820	20.959
		0.3884	0.0004	0.11		

The three values in the bottom row represent the average, standard deviation, and relative standard deviation for the calculated value, which is in this case, the anisotropy, r. So,

$$r_{\text{avg}} = 0.3884, s = 0.0004, \text{rsd} = s / r_{\text{avg}} = 0.11$$

### Set 3: Polarization data collected for one sample with 3 replicates.

## Advanced Reads Report

Report Time : Thu 18 Apr 05:25:06 PM 2002  
Software Version: 1.0(83)  
Operator:

### Instrument Parameters

Instrument Cary Eclipse  
Instrument Serial Number e100083670  
Data mode Fluorescence  
User Result execute("AutoPolarizationCollect.ADL")  
Ex. Slit (nm) 5  
Em. Slit (nm) 5  
Ave Time (sec) 5.0000  
Excitation filter Auto  
Emission filter Auto  
PMT Voltage (V) Medium  
Replicates 3  
Sample averaging OFF  
Comments:

### G-Factor

Instrument 5  
Data mode Fluorescence  
Ex. Slit (nm) 5  
Em. slit (nm) 5  
Ave. time(s) 5.00000

Ex. WL (nm)	Em. WL (nm)	G-Factor	Int(HV) (a.u)	Int(HH) (a.u.)
555.00	575.00	2.0753	623.954	300.651

### Analysis

Collection time 4/18/02 5:25:23 PM

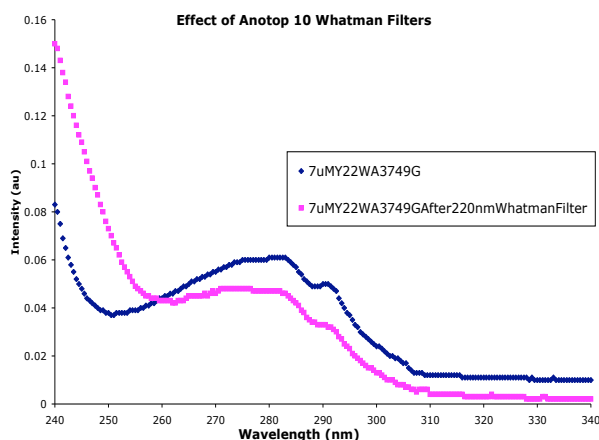
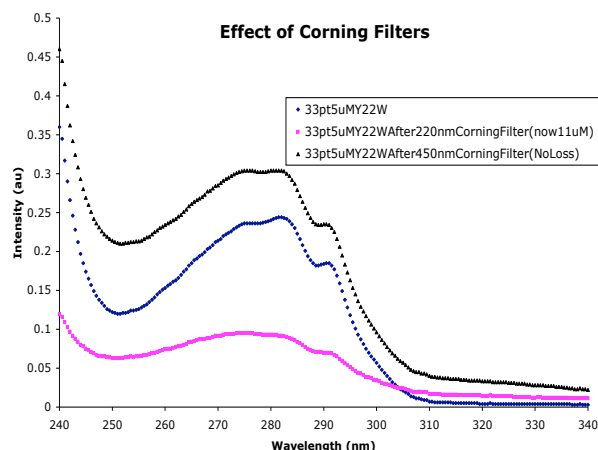
### Polarization

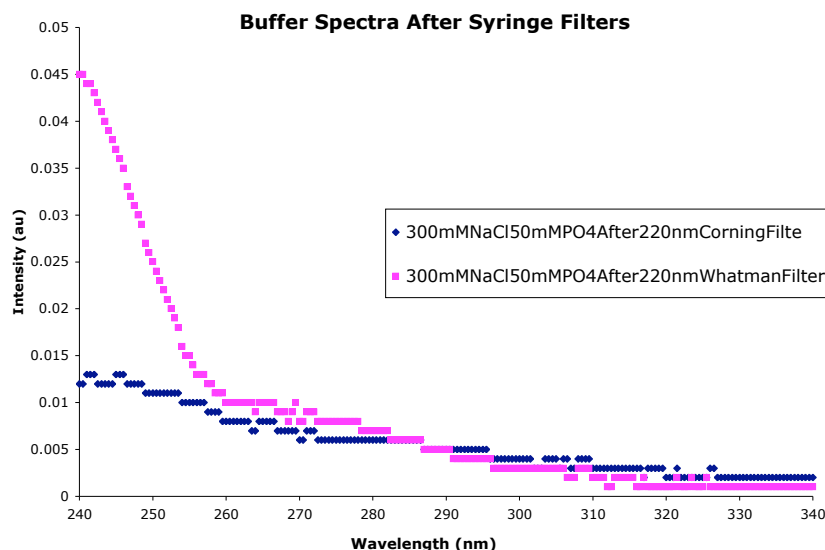
Sample Name	Ex. WL (nm)	Em. WL (nm)	p	G-Factor	Int(VV)	Int(VH)
Sample 1	555.00	575.00	0.49	2.0753	125.960	20.960
Sample 1	555.00	575.00	0.49	2.0753	125.837	21.000
Sample 1	555.00	575.00	0.49	2.0753	126.128	21.014
		0.4861	0.0006	0.11		

Again, the single point measurements for polarization and anisotropy compare well with the expected values for a fluorophore in a rigid environment that prevents rotation of the molecule, which results in minimal depolarization of the light following excitation.

## A.8 Proper Use of Syringe Filters

It was noted on 5-July-2007 that the membranes used for syringe filtration of proteins in the lab were either leaching far UV-absorbing chemicals into the filtrate (this was observed for both buffer and protein solutions-the Whatman filter released the largest amount of impurities), binding large amounts of protein so that recovery was very low (only ~33% for the Corning filter, very low yield-but not quantifiable-for the Whatman filter due to UV-absorbing impurities), and denaturing proteins (this was evident from the distortion of the ~280nm protein peak in the filtrate from all three filters). Overall, the worst offender (with very low recovery, denaturant activity, and UV-absorbing impurity leaching) is the 220nm Whatman Annotop 10 filters. The 220nm Corning SFCA (surfactant-free cellulose acetate) filter was not quite as bad (recovery was ~33%, impurity leaching was negligible, UV absorbance still revealed denatured protein), and the Corning 450nm SFCA filter was the best of the three (there is still some amount of denaturing, but recovery is ~100% and leaching is negligible). The above effects were observed even after copious pre-rinsing with buffer.





Manufacturer notes on these filters:

### 1. Whatman Anotop® syringe filters

Anotop disposable syringe filters are designed for use with most organic solvents and aqueous materials and are suitable for sample volumes up to 100 mL. The devices feature a distinctive hexagonal housing, produced from pigment-free polypropylene to eliminate sample contamination. No wetting agents or adhesives are used in the manufacturing process.

Typical Data - Anotop Syringe Filters				
	Anotop 10	Anotop 10 Plus	Anotop 25	Anotop 25 Plus
Housing	Polypropylene	Polypropylene	Polypropylene	Polypropylene
Filtration Area	0.78 cm <sup>2</sup>	0.78 cm <sup>2</sup>	4.78 cm <sup>2</sup>	4.78 cm <sup>2</sup>
Maximum Pressure	100 psi	100 psi	100 psi	100 psi
Volume 'Hold Up'	<20 µL	<30 µL	<150 µL	<200 µL
Prefilter Type	N/A	Glass microfiber (binderless)	N/A	Glass microfiber (binderless)
Membrane Diameter	10 mm	10 mm	25 mm	25 mm
Membrane Type	Anopore	Anopore	Anopore	Anopore
Average Membrane Thickness	60 µm	60 µm	60 µm	60 µm
Device Width	14 mm	14 mm	31 mm	31 mm
Device Length	18 mm	18 mm	25 mm	25 mm
Device Shape	Hexagonal	Hexagonal	Hexagonal	Hexagonal
Construction Process	Thermal weld	Thermal weld	Thermal weld	Thermal weld
Inlet Connection	Female luer lock	Female luer lock	Female luer lock	Female luer lock
Outlet Connection	Male slip luer	Male slip luer	Male slip luer	Male slip luer
Protein Adsorption	Low	Medium/high	Low	Medium/high
Extractable Materials	Low	Low	Low	Low
Cytotoxicity	Non-cytotoxic	Non-cytotoxic	Non-cytotoxic	Non-cytotoxic

## Anotop 10 Features and Benefits

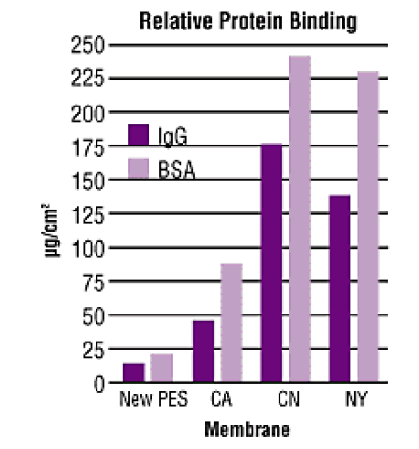
- \* 10 mm diameter syringe filter
- \* Sample volume up to 10 mL
- \* Inorganic membrane
- \* Capillary pore structure
- \* Made from Gamma-Alumina 6 mm Al<sub>2</sub>O<sub>3</sub>
- \* Low protein binding
- \* Low hold up volume <20  $\mu$ L ensures maximum sample recovery
- \* Sterile formats available for critical applications

2. Corning SFCA Filters: Cellulose acetate (CA) membranes have a very low binding affinity for most macromolecules and are especially recommended for applications requiring low protein binding, such as filtering culture media containing sera. However, both cellulose acetate and cellulose nitrate membranes are naturally hydrophobic and have small amounts (less than 1%) of non-toxic wetting agents added during manufacture to ensure proper wetting of the membrane. If desired, these agents can be easily removed prior to use by filtering a small amount of warm purified water through the membrane or filter unit. Surfactant free cellulose acetate membranes, with very low levels of extractables, are available on some Corning syringe filters.

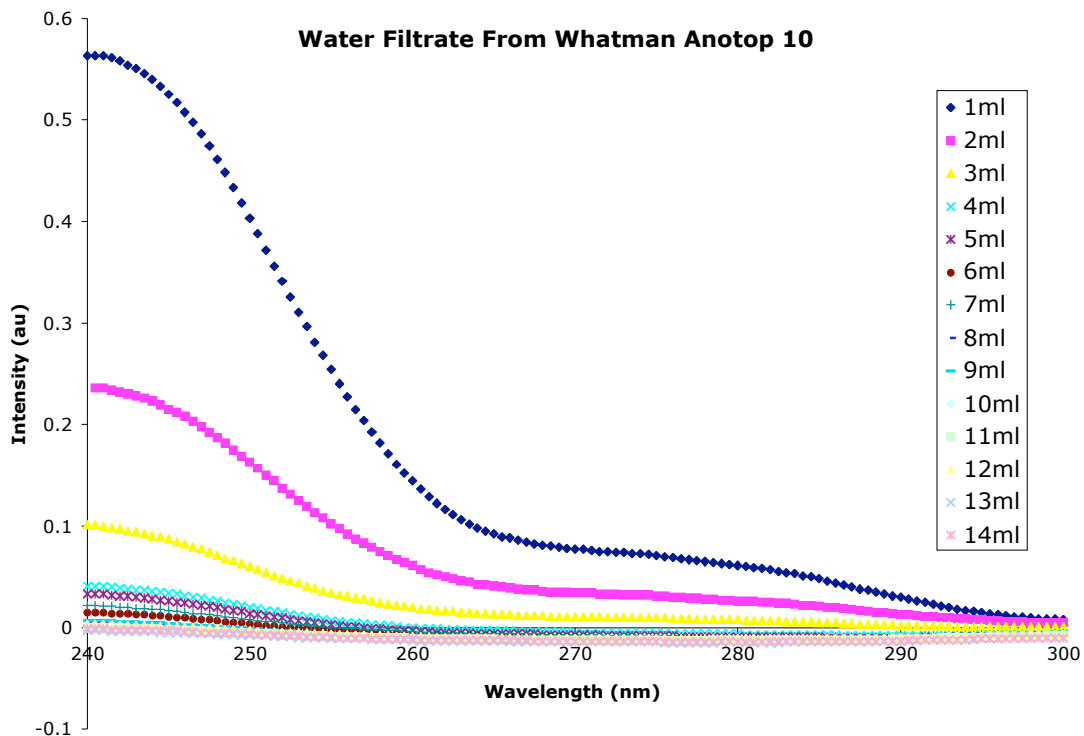
**Table 2. Characteristics of Corning Filter Membranes**

	Cellulose Nitrate	Cellulose Acetate	Nylon	Polyether-sulfone	Regenerated cellulose	Teflon (PTFE)
Wetting Agents	Yes	Yes	No, Naturally Hydrophilic	No	Yes	Does not wet
Protein Binding	Very High	Very Low	Low to Moderate	Very Low	Low	NA
DNA Binding	High	Very Low	Very High	Very Low	Low	NA
Chemical Resistance	Low	Low	Moderate to High	Low	Very high	Very high

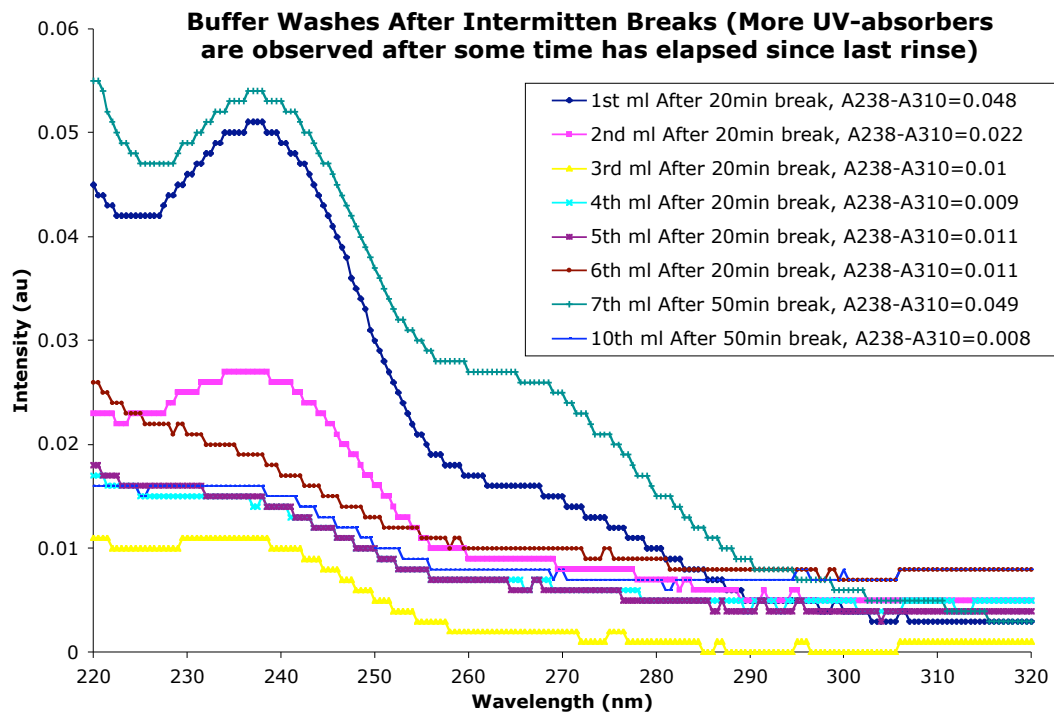
Other potential filter material: Polyethersulfone (PES) membranes are highly recommended for filtering cell culture media. PES has both very low protein binding and extractables. PES also demonstrates faster flow rates than cellulosic or nylon membranes.



Trial 1: Can UV-absorbing material leeching out of Whatman Anotop 10 filters be removed by rinsing with water?



Trial 2: What is the effect of leaving filters (wet at room temperature) for 10-30 minutes after rinsing?





## **Appendix B**

### **Materials for the UIUC Biophysics K-12 Outreach Program**

#### **B.1 Program Overview**

Our outreach program pairs biophysics graduate students with local K-12 teachers to develop lessons that teach middle and high school students about the exciting research happening here at the University of Illinois, Urbana-Champaign.

The teachers we work with come from local and suburban Chicago schools where they teach physics, biology or chemistry to students from 7th to 12th grade.

The graduate students are PhD candidates doing research on a vast array of topics in biophysics, from single molecule fluorescence microscopy to molecular dynamics simulation of membrane proteins. These students come from diverse academic backgrounds, including physics, biology and computer science and are eager to give back to the community while improving their ability to teach and communication their research.

#### **B.2 Biophysics Outreach Manual For Graduate Students**

*For more information please see our website:*

*<http://www.life.uiuc.edu/biophysics/ibs/K-12Outreach/Homepage.html>*

Center for Biophysics and Computational Biology

University of Illinois, Urbana-Champaign

607 South Mathews Avenue Urbana, IL 61801 USA

Ph: (217) 333 1630, FAX: (217) 244 6615, E-Mail: [biophysics@life.uiuc.edu](mailto:biophysics@life.uiuc.edu)

Director: Robert M. Clegg    Administrative Coordinator: Cindy Dodds



*“...in addition to being competent researchers, STEM graduate students must be able to communicate science and research to a variety of audiences. As the graduate students bring their cutting-edge research and practice into the K-12 classroom, they gain these skills which enable them to explain science to people of all ages, ranging from students to teachers. The graduate students also inspire transformation in the K-12 formal and informal learning environments and stimulate interest in science and engineering among students and teachers.”*

- National Science Foundation, GK-12 Program website, [www.nsfgk12.org](http://www.nsfgk12.org)

## **Introduction**

Scientists have been participating in education outreach since at least 1820, when Michael Faraday began his famous Christmas Lectures. More recently though, due to sobering studies exposing the growing science and math deficiencies among American

children, there has been a growing number of scientists participating directly in K-12 classrooms. These outreach programs have often proved successful in significantly improving the level of inquiry and enthusiasm for science among both teachers and students. In 1999, the National Science Foundation created a new graduate fellowship, the NSF GK-12 Program, that pairs science PhD students with K-12 teachers in an effort to improve communication and teaching skills among graduate students while enriching science and math instruction for their K-12 partners. Since its inception, the GK-12 program has funded more than 200 projects at universities across the US, taking graduate student outreach in K-12 classrooms to an all time high.

The UIUC GK-12 program has been particularly successful in bringing UIUC graduate students into K-12 classrooms. Sadly, this program will come to an end by summer 2009, leaving behind a great demand for partnerships between university researchers and local schools. We hope our program will help to continue the fantastic work of the GK-12 program with biophysics graduate students whose interdisciplinary research can be taught in a wide range of K-12 classrooms.

## **How the Program Works**

Students interested in participating in this program should first contact the Program Director, Bob Clegg to discuss their lesson ideas and have them approved. This pre-approval will be necessary in order to receive tutorial credit for the outreach. In looking for lesson ideas, your research advisor is also a great source of information. We highly encourage research advisors to become involved in this process.

In order to have sufficient time for lesson development, it is a good idea to begin this process at least one month prior to the day(s) you want to teach your lesson. You may need an additional 1-2 weeks afterward to write up your experience for tutorial credit.

Based on your lesson, you will be directed to one or more teachers from the list below (also available on our website). It is important that the teacher's subject and grade level is appropriate for the topics you would like to teach. When you contact a teacher, briefly describe one or more ideas you have for a lesson and the dates you are available to teach in the classroom. You will get the best results if you are flexible in terms of dates and times since most teachers follow a pretty strict curriculum that may be amenable to your lesson only at certain times of the year. The teacher will respond letting you know if they can or cannot accommodate your lesson and may suggest some initial modifications based on time, space or curricular constraints of the classroom.



Once you have found a partner teacher and have scheduled a date for your lesson, you will need to schedule a time to meet with the teacher by phone or in person as soon as possible. Your initial meeting should be at least 3 weeks before your lesson and you will need to have a rough draft of the lesson plan for you and the teacher to review. The teacher will suggest more changes to improve the accessibility of language and content to the students and to ensure that the lesson is doable. You will want to pay very close attention to ensure your final draft reflects the teacher's suggestions. You will want to ask the teacher what students will have already learned about your subject area, how many students there are, how much space is available, what type of equipment, materials, and resources the classroom and school can provide, and how much time is actually allotted for teaching (versus roll call, announcements, etc.). It is VERY IMPORTANT to

ensure you and the teacher are in agreement about what materials and equipment they will provide and what you will need to bring. This includes everything from lab chemicals and glassware, to photocopies, computers, projectors and even chalk. Knowing who is responsible for what will ensure your lesson runs smoothly.

Based on this initial meeting, you will revise your lesson plan and complete a materials list which details who will be providing what equipment or materials and a student handout, if any. Try to make your lesson plan as detailed as possible in case future graduate students or teachers want to use it. Check the link on the website for sample lesson plans. The most successful outreach lessons are ones that contain a hands-on activity, lab or an engaging demonstration. Your lesson should not just tell students about scientific discoveries, but allow them to experience the joy of those discoveries for themselves.

Email your revised lesson plan to the teacher at least two weeks prior to the lesson date. The teacher may provide final feedback and you will need to get your final version to them at least one week before the lesson.

### **A note on materials and equipment:**

If there are any materials that the school cannot provide for your lesson, first check the inventory of the Illinois Biophysics Society's outreach program. There are also on-campus science outreach programs and teaching labs (see links below and on our website) that often have materials and equipment you can borrow. Many items can also be donated or borrowed from research labs. Liquid nitrogen, for example, is only \$0.2/Liter and often labs have equipment, such as a vacuum pump and dessicator that they can live without for a day. If you are still not able to find the materials you need, contact the Biophysics office, as there is a small budget for purchasing new outreach materials.

## Summary of Program Structure & Timeline:

- Read through this manual and reference 1 below. Come up with a couple ideas for lessons based on your research and make an appointment with the Director, Bob Clegg.
- Once you're approved, e-mail teachers with your ideas and available dates (dates



should be at least 3 weeks away).

- Arrange to meet in person or by phone with your partner teacher at least 3 weeks before your lesson date. Go over all details of your lesson including all time, curricular, material and space constraints of the classroom. Be clear about where all equipment and materials are coming from and what changes the teacher wants. you to make to ensure the scope and difficulty level is appropriate.
- Revise your lesson plan and email/send to your partner teacher for final comments at least 2 weeks before your lesson date.
- Check with IBS outreach, on-campus outreach programs, on-campus teaching labs and research labs for equipment and materials. Contact the Biophysics office ASAP if you need to request purchase of new materials.
- Send the final draft of your lesson plan to your partner teacher at least 1 week before the lesson date.
- Have fun!

## General Advice:

- Remember to promote inquiry whenever possible through the types of questions you ask (open-ended), the design of your labs (not cookbook recipes, but labs with a different result every time). You can also achieve this by using hands-on activities (allow students to experience discovery for themselves) and by ensuring the language and content of your lectures, handouts, or other reading material is accessible to students.
- Be concise in your correspondence with teachers; they don't have a lot of time for email.
- Be responsible with borrowed equipment! If you break something it will jeopardize relationships and resources for future volunteers.
- Allow several weeks for any new software to be installed on school computers. Also, check them in person before the day of your lesson.
- Try not to change dates or cancel once you've already committed with a teacher. Doing this could jeopardize relationships for future volunteers.
- Check the [Illinois State Board of Education State Standards](http://www.life.uiuc.edu/biophysics/ibs/K-12Outreach/Homepage.html) to help you determine whether your lesson ideas will fit into the curriculum and at which age/level.



## B.3 Biophysics Outreach Manual For Teachers

*For more information please see our website:*

*<http://www.life.uiuc.edu/biophysics/ibs/K-12Outreach/Homepage.html>*



Center for Biophysics and Computational Biology  
University of Illinois, Urbana-Champaign  
607 South Mathews Avenue Urbana, IL 61801 USA  
Ph: (217) 333 1630, FAX: (217) 244 6615, E-Mail: [biophysics@life.uiuc.edu](mailto:biophysics@life.uiuc.edu)  
Director: Robert M Clegg    Administrative Coordinator: Cindy Dodds



*“...in addition to being competent researchers, STEM graduate students must be able to communicate science and research to a variety of audiences. As the graduate students bring their cutting-edge research and practice into the K-12 classroom, they gain these skills which enable them to explain science to people of all ages, ranging from students to teachers. The graduate students also inspire transformation in the K-12 formal and informal learning environments and stimulate interest in science and engineering among students and teachers.”*

- National Science Foundation, GK-12 Program website, [www.nsfgk12.org](http://www.nsfgk12.org)



## Introduction

Scientists have been participating in education outreach since at least 1820, when Michael Faraday began his famous Christmas Lectures. More recently though, due to sobering studies exposing the growing science and math deficiencies among American children, there has been a growing number of scientists participating directly in K-12 classrooms. These outreach programs have often proved successful in significantly improving the level of inquiry and enthusiasm for science among both teachers and students. In 1999, the National Science Foundation created a new graduate fellowship,

the NSF GK-12 Program, that pairs science PhD students with K-12 teachers in an effort to improve communication and teaching skills among graduate students while enriching science and math instruction for their K-12 partners. Since its inception, the GK-12



program has funded more than 200 projects at universities across the US, taking graduate student outreach in K-12 classrooms to an all time high.

The UIUC GK-12 program has been particularly successful in bringing UIUC graduate students into K-12 classrooms. Sadly, this program will come to an end by summer 2009, leaving behind a great demand for partnerships between university researchers and local schools. We hope our program will help to continue the fantastic work of the GK-12 program with biophysics graduate students whose interdisciplinary research can be taught in a wide range of K-12 classrooms.

## How the Program Works

Teachers interested in participating in this program should contact the Director, [Bob Clegg](#), with their name, school, classes they would like to have participate in the program, and their contact information. Before working with a graduate student, it is advisable that teachers read through this brief manual, especially if you have not worked with research scientists in your classroom before.

When a UIUC biophysics graduate student has a lesson that they think will be appropriate for your students, they will send you an e-mail with a brief description of what they would like to do along with a range of dates when they are available to do it. You are not required to accept requests from students simply because you are part of the program. Rather,

you should decide whether the proposed lesson (or a slight variation of it) will enrich your course, and whether you can fit it in during a time the graduate student is free. If you can agree on a date,



great! Ideally this date will be at least 4 months away to allow sufficient time for lesson development and revision.

The next step is to arrange a meeting time over the phone or in person with the graduate student. The graduate student will bring a rough draft of the lesson to this meeting, including any student handouts or other materials to be printed. During this meeting, it is crucial that you offer your expertise regarding the accessibility of the

language and content of the lesson to your students. Graduate researchers will have little to no experience communicating any science, let alone sophisticated research, to K-12 audiences. One of the goals of the lesson is for the graduate student to be able to communicate some aspect of their research to the students and to allow the students to experience the joy of discovery themselves through a hands-on activity, demonstration or lab experiment.

It is particularly important at this initial meeting to assess the feasibility of the lesson plan given the time, space and curricular constraints of your classes. Please be as clear as you can with the graduate student regarding what materials and equipment you can provide, from photocopies to lab glassware, computers, projectors, etc. By the end of your first meeting, you and the graduate student should be perfectly clear about who is providing what materials on the day of the lesson.

It will also be helpful if you can tell the graduate student what your students will have learned about the lesson topic by the time it is taught. You will also need to provide the graduate student with several logistical details such as your class size, the time allotted for actual teaching (after roll call, announcements, etc.) and any special school rules you think he or she will need to know.



Based on this initial meeting, the graduate student will revise the lesson plan and complete a list detailing who will be providing what equipment or materials. This semi-

final draft will be e-mailed to you about 2-weeks before the lesson date. At this point, it is crucial that you read the lesson over and ensure it is feasible and appropriate for your students. Any feedback you can provide the graduate student will be very helpful for them.

The graduate student will then make the final changes and e-mail you the final draft one week before the lesson date. If there are any changes in the materials you can provide, please let the graduate student know as soon as possible so they can look for alternate sources.



## **B.4 Other outreach programs**

The references at the end of this manual will help you familiarize yourself with current issues in science education reform, but perhaps the best information sources are the websites of the following K-12 outreach programs:

### **UIUC:**

- [Chemistry Outreach](#)
- [The Physics Van](#)
- [G.A.M.E.S.](#)
- [Engineering Outreach](#)

### **Illinois:**

- [UIC Weird Science](#)
- [MSCOPE \(University of Chicago\)](#)
- [Museum of Science & Industry](#)
- [Orpheum Children's Science Museum](#)

### **US:**

- [Caltech Precollege Science Initiative](#)

- [NSF GK-12 Programs](#)
- [SEPAL \(San Francisco State\)](#)
- [Exploratorium Hands on Activities](#)

## **B.5 Sample lessons**

There are several interesting sample lessons in the articles in the references section below, but the best examples will be those provided by previous UIUC biophysics student volunteers and their partner teachers. These can be downloaded from the program [website](#). At this point, there has been only one completed outreach lesson, which is described below.

### **B.5.1 Introduction to Structural and Molecular Virology by Ross Bodnar**

In the spring of 2009, Ross Bodnar worked together with Danville High School biology teacher Kathy Hafner to develop and teach a lesson on the topic of Molecular and Structural Virology. Ross prepared a brief lecture before leading students through explorations of viral architecture using the web-based software “Virus Particle ExploreR<sup>2</sup>” which employs the database of icosahedral virus capsid structures known as Viperdb. These resources are maintained by The Scripps Research Institute at <http://viperdb.scripps.edu/>. The Power Point lecture slides and student handout for this lesson are available on our program website.

## **B.6 Science Education Journals**

Use the following journals to search for sample lessons or for information about current issues in science education reform and best teaching practices for K-20 science and engineering classrooms. Should you decide to publish your outreach lesson(s) and

any corresponding assessment results, the first five journals would be the most appropriate places to send your work.

- [The American Biology Teacher](#)
- [The Science Teacher](#)
- [The Journal of Biological Education](#)
- [The Chemical Educator](#)
- [The Physics Teacher](#)
- [The Journal of Chemical Education](#)
- [Chemical Engineering Education](#)
- [Life Sciences Education](#) (*formerly Cell Biology Education*)
- [Electronic Journal of Science Education](#)
- [International Journal of Science Education](#)
- [Journal of Science and Technology Education](#)
- [Journal of Engineering Education](#)
- [Journal of Technology Education](#)
- [Journal of Research in Science Teaching](#)

## **B.7 Directory of participating schools and teachers**

Because Biophysics graduate students have very diverse interests and expertise, the partner teachers for our program come from a variety of subject areas and grade levels. Graduate students should choose a classroom that best matches their research interests. Teachers who have participated in outreach with graduate students in the past are marked with an asterisk. They will be the best prepared to help you integrate your research into an age-appropriate lesson.

<u>School/Address</u>	<u>Teacher/Admin &amp; Subject(s)</u>
<a href="#">University Lab High School</a> 1212 W. Springfield Ave. Urbana, IL 61801 (217) 333-2870	<a href="#">David R. Bergandine</a> , Chemistry Teacher
<a href="#">Urbana Middle School</a> 1201 South Vine St. Urbana, IL 61801 (217) 384-3685	* <a href="#">Jan Hari</a> , Science Department Coordinator & 8 <sup>th</sup> grade Science Teacher
<a href="#">Edison Middle School</a> 306 W Green St. Champaign, IL 61820 (217) 351-3771	* <a href="#">Kevin Kuppler</a> , 7th grade Science Teacher * <a href="#">Bryan Foli</a> , 7th grade Science Teacher
<a href="#">Danville High School</a> 202 East Fairchild St. Danville, IL 61832 (217) 444-1619	* <a href="#">Kathy Hafner</a> , Biology (honors & AP) Teacher <a href="#">Beth Chamberlain</a> , Co-Division Leader & Physics (merit & Engineering) Teacher
<a href="#">Neuqua Valley High School</a> 2360 95th Street Naperville, IL 60654 (630) 428-6040	<a href="#">Herb Anderson</a> , Physics Teacher <a href="#">Mark Cummings</a> , Biology (honors) Teacher <a href="#">Patti Smykal</a> , Chemistry (honors and AP) Teacher <a href="#">Lynn Vlcek</a> , Biology (AP) & Chemistry (AP) Teacher <a href="#">Jaki Mimnaugh</a> , Genetics & Physical Science Teacher
<a href="#">Metea Valley High School</a> (Opening Fall 2009) N. Eola Road Naperville, IL (708) 925-1357 (cell)	* <a href="#">Matt Kirkpatrick</a> , Science Department Chair, (honors & AP Biology & Chemistry teachers are available to participate)



## B.8 References

These papers provide some insight into current issues in science education as well as some interesting sample lessons.

1. Bower, James. M. (1996). Scientists and science education reform: myths, methods, and madness. The National Academies RISE (Resources for Involving Scientists in Education) from <http://www.nas.edu/rise/backg2a.htm>.
2. RamosOrtiz, G., Maldonado, J. L., Villagran-Muniz, M., and Ramirez-Maldonado, Z. (2000). Forty simple experiments with an He-Ne laser for high school students. Sixth International Conference on Education and Training in Optics and Photonics, 101(82), 282-292.
3. Tomanek, D. (2005). Building Successful Partnerships Between K–12 and Universities. *Cell Biol. Educ.*, 4, 28-37.
4. Elgin, S.C.R., Flowers, S., May, V. (2005). Modern genetics for all students: an example of a high school/university partnership. *Cell Biol. Educ.*, 4, 32-34.
5. Dolan, E., and Tanner, K. (2005). Moving from outreach to partnership: striving for articulation and reform across the K–20+ science education continuum. *Cell Biol. Educ.*, 4, 35-37.
6. Schultz, T. (1996) Science education through the eyes of a physicist. The National Academies RISE (Resources for Involving Scientists in Education) from <http://www.nas.edu/rise/backg2a.htm>.
7. White, B., Kim, S., Sherman, K. and Weber, N. (2002) Evaluation of molecular visualization software for teaching protein structure. Differing outcomes from lecture and lab. *Biochemistry and Molecular Biology Education*, 30(2), 130-136.



8. Ricks, I. (2004) The 50th Anniversary of Brown v. Board of Education: Continued impacts on minority life science education. *Cell Biol. Educ.*, 3, 146-149.
9. Nelson, A. and Goetze, J. (2004) Modeling protein folding and applying it to a relevant activity. *The American Biology Teacher*, 66(4), 287-289.
10. Making holograms In middle and high schools. Sixth International Conference on Education and Training in Optics and Photonics, 101(82), 223-228.
11. Barnard, B. Microarrays made simple: “DNA chips” paper activity. *The American Biology Teacher Online Publication*, March, 2006, 8-12.
12. Seipelt, R. Cookie-ases: Interactive Models for Teaching Genotype-Phenotype Relationships. *American Biology Teacher Online Publication*, March, 2006, 8-12.
13. Puterbaugh, M.N. & Burleigh, J.G. (2001) Investigating evolutionary questions using online molecular databases” *The American Biology Teacher*, 63(6), 422-431.
14. Seelman, G. A scientist’s guide to making successful presentations to high school students. *The National Human Genome Project*, <http://genome.gov/Education>. Accessed April 6th, 2009.
15. Ledley, T., Haddad, N., Lockwood, J., Brooks, D. (2003) Developing Meaningful Student-Teacher-Scientist Partnerships. *Journal of Geoscience Education*, 5(1), 91-95.

16. Bilbro, J., Walker, J. G. (2000) Optics education for K-12. Sixth International Conference on Education and Training in Optics and Photonics, 101(82), 92-97.
17. Hall-Wallace, M., Regans, N.L. (2003) Impact of a K-12 Partnership on Science Teaching. *Journal of Geoscience Education*, 51(1), 104-113.

## Appendix C

### Molecular Visualization and Bioinformatics Lesson Materials

#### C.1 How to Download VMD

1. You can download VMD from the following website:  
<http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD>
2. For windows machines, choose the Windows OpenGL download option.
3. You will first need to register before you can download the software. Enter in your first name and last initial (no spaces) as your user name. Make up a 6 character password and press “continue”.
4. Enter in your first and last name & your email address. Your affiliation is “Academic”, the number of people using the software is “1”, & you are using the software for “teaching”.
5. You are not funded by NIH, so select “no” for this question. Now, re-enter your password and select “I Agree to the Terms of this License”.
6. Now the download should begin. You will be asked to save the VMD installer; save it to the desktop.
7. Minimize your browser window and double click on the icon labeled “VMD 1.6”. Click “setup” to extract the installer. The installer wizard will automatically load after extraction. Select “next”, then “yes” (to the license agreement), then “next” 3 more times & the files should begin installing. When it is done, click finish on the installer wizard.

### *How to Obtain a Structure From the Protein Data Bank*

1. The easiest way to do this is to simply open up VMD, select “New Molecule” from the file menu in the VMD Main window, and type in your protein’s PDB ID followed by a carriage return (hit the enter key on your keyboard). The file type drop down menu should automatically change to “Web PDB Download”. If it does not, you should change this yourself. Now, just make sure you are connected to the internet and press the “load” button. It may take a few minutes for your protein to load, so be patient!
2. If you don’t know your protein’s PDB ID, then you can go to the Protein Data Bank site, [www.rcsb.org](http://www.rcsb.org), to get it. Scroll to the bottom of the page to where it says “Molecule of the Month”. Click on the “Previous Features” link. Look for your protein in this list and click on its name. Links to the PDB IDs for structures of your protein will be located in these description pages. If there are multiple PDB IDs listed, choose the one on the last page titled “Exploring the Structure”. Now that you have the PDB ID, go back and follow the instructions in step 1!

## **C.2 What’s So Special About Water**

### **C.2.1 Lesson Plan**

#### *Lesson Goals:*

1. Students will have an improved understanding of how hydrogen bonds give rise to most of water’s life-sustaining properties: its high surface tension, heat capacity,

enthalpy of vaporization, melting and boiling points, the expansion of water upon freezing, and water's ability to act as a "universal solvent".

2. Students will understand why the above properties are important for life on earth.
3. To introduce students to computer molecular visualizations and simulations.

*Prerequisite Knowledge:*

A basic knowledge of the molecular properties of water, its chemical formula, polarity, geometry and its ability to be both a donor and acceptor of hydrogen bonds.

*Materials:*

- "What's So Special About Water" Power Point presentation (Water.ppt).
- One computer with VMD installed, an LCD projector and a screen.
- Simulation files: Melting.psf and Melting.dcd

*Protocol:*

1. Begin class with the Power Point presentation covering the structure of water, hydrogen bonding and how hydrogen bonding give rise to the special properties of water that are crucial for life to exist on earth.
2. Load the simulation files by clicking "New Molecule" from the File menu on the VMD Main window and browsing to one of the .psf files (either Melting.psf or SurfaceTension.psf). Click "Load" and then highlight the molecule name in VMD Main before selecting "Load Data into Molecule" from the File menu.

Now browse to find the .dcd file corresponding to the .psf file you chose and click “Load”.

3. The simulation should immediately play, but if you want to loop the simulation, stop it, change the step size or pause, use the controls on the bottom of the VMD Main window.
4. The simulations are best viewed with the CPK drawing method. To change the drawing or coloring method, choose “Representations” from the Graphics tab in the VMD Main window. Under drawing method, click CPK. You can play around with other drawing and coloring methods as well.
5. Engage the class with questions about the simulation. Explain the CPK coloring system and then ask if they can see the individual water molecules. Can they tell which are hydrogen bonded to each other? Can they tell when the ice begins to melt? Can they guess why the water box deforms into a sphere (to minimize the surface tension, since it is surrounded by vacuum)?

*Grading/assessment:*

Teachers may give points for class participation or include discussion questions on the next quiz or exam.

### **C.3 Visualizing Biopolymers and Their Building Blocks**

### C.3.1 Lesson Plan

#### *Lesson Goals:*

1. To improve students' understanding of the molecular structure of biopolymers and their building blocks.
2. To improve retention of molecular knowledge by providing problem solving opportunities in a unit dominated by rote memorization.

#### *Prior knowledge required:*

- Cells make a huge number of large molecules called polymers from a limited set of small molecules called monomers. When the monomers are identical, as with starch, we call it a homo-polymer. If there are two or more different monomer types, as with proteins, it is a hetero-polymer.
- Polymers are formed in a condensation reaction (dehydration synthesis) in which two monomers lose a water molecule and are joined covalently. Hydrolysis, the opposite of dehydration synthesis, is when the covalent bonds linking two monomers in the polymer are broken and water is gained.
- Biopolymers fall into four distinct classes: proteins, nucleic acids, lipids, and carbohydrates, which are made up of amino acids, nucleosides, fatty acids and monosaccharides.

#### *Materials:*

- A computer with VMD installed and an LCD projector for the instructor.

- A computer with VMD installed for each student group.
- Student handout (OrganicMoleculesVMD.pdf)
- Molecule files:
  - a. Carbohydrates: glucose.pdb, fructose.pdb, sucrose.pdb, starch.pdb
  - b. Lipids: palmitic.pdb, oleic.pdb, glycerol.pdb, popc.pdb, cholesterol.pdb
  - c. Nucleic Acids: thymine.pdb, uracil.pdb, adenosine.pdb, dna.pdb
  - d. Proteins: glycine.pdb, dipeptide.pdb, helix.pdb, betasheet.pdb

*Protocol:*

*1-2 Days before*

Prepare students for the activity by going over relevant terms and concepts in the “Prior Knowledge Required” section above.

*Day 1*

Lead students through the “Loading a Molecule in VMD” section of the handout using a molecule of your choice. Take the time to remind students of the CPK coloring conventions and to teach them the shortcuts for basic molecule manipulations in VMD. Lead the class through the first question in the chosen section (“Carbohydrates” if it is done in order). Students can complete the rest of the activity on their own or in groups of 2-3.

*Days 2 - 4*

Begin class by a brief refresher on how to load and manipulate molecules in VMD. Lead the class through the first question in the subsequent section. Students should finish one section each class period.



Note to the instructor:

One or more sections can be skipped if there is limited time. Each section requires one full class period. Ball and stick models are a very useful augmentation to these activities if they are available.

*Grading/assessment:*

If students work in teams, 12 points are given for teamwork and 22 points are equally divided among the questions in each section. If students work individually, 25 points are given per section. The extra credit in the “Lipids” section is worth 5-10 points.

### C.3.2 Student Handout

## Visualizing Biopolymers and Their Building Blocks

Living things are primarily composed of carbon-based (organic) polymers. These are made up many small molecules which have been chemically bonded together to form a long chain. Biopolymers can be grouped into 4 main categories: LIPIDS, CARBOHYDRATES, PROTEINS, & NUCLEIC ACIDS. In this exercise, you will use the program VMD to explore the structures and learn more about these molecules and their building blocks.

### Loading a Molecule into VMD

1. Click on the “Start” menu, then go to “programs” and then to “VMD 1.8.3”. The window labeled “VMD Main” is the control panel and the window labeled “OpenGL” is where the molecule(s) will be displayed.
2. In the VMD Main window, select “file”, then “new molecule”. Hit the “Browse” button to search for the file you want. They will be stored under your class folder on the “Science on DHS Users” server. Once you select the file you want, you will hit load. The molecule you want should appear in the OpenGL window. If it does not, please see the teacher for help.
3. Once your molecule is loaded, go to the VMD Main menu & select “Representations” under the Graphics menu. On the bottom panel, change the “Drawing Method” to “CPK” to better visualize the molecule.

4. You can use your mouse to manipulate the structure by using the keyboard shortcut commands described below.
5. When you are all finished, go to the VMD main menu and select “molecule” and then “delete molecule” before loading the next structure into VMD.

**List of VMD Keyboard Shortcuts (also accessible under the “mouse” menu):**

<b>Key pressed</b>	<b>Action Performed</b>
<b>t</b>	Translate, or move entire molecule around
<b>r</b>	Rotate molecule around its center
<b>s</b>	Make the molecule bigger (mouse to left) or smaller (mouse to right)
<b>C</b>	You can select “c” and then click an atom which you would like your structure to rotate around
<b>X</b>	Rotate your molecule about the x axis
<b>Y</b>	Rotate your molecule about the y axis
<b>Z</b>	Rotate your molecule about the z axis
<b>1</b>	Label atoms selected with mouse left click
<b>2</b>	Enters “bond label” mode, which gives the distance between two atoms selected by successive mouse left clicks

## CARBOHYDRATES

The basic building block of carbohydrates are sugar monomers, or monosaccharides. Carbohydrates include a wide variety of important and diverse molecules from the starch which is the ultimate food source for most of the biosphere, to cellulose, which is an essential structural component of plant cell walls. In this assignment, you will explore the structure of 2 common monosaccharides as well as one disaccharide, and small starch molecule.

1. First you will explore the structure of glucose. How many carbons, hydrogens and oxygens are present? Write down the chemical formula here:

Draw a 2-dimensional representation of the glucose structure in the space below:

2. Delete the glucose molecule and load the fructose molecule. How many carbons, hydrogens and oxygens are present? Write down the chemical formula here:

Compare the chemical formula of glucose and fructose. What is the special relationship between these two sugars (if you can't remember the name, just describe it in your own words)?

All sugars have the same ratio of carbon to hydrogen to oxygen. Write down that ratio here:

Draw a 2-dimensional representation of the fructose structure in the space below:

3. Delete the fructose molecule and load the sucrose molecule. How many carbons, hydrogens and oxygens are present? Write down the chemical formula here:

Sucrose is a disaccharide. This means that it is made up of two sugar monomers or monosaccharides. Are the two sugars that make up sucrose the same or different?

Write the sequence of sucrose below (for example, maltose is a homo-disaccharide of glucose, so its sequence is **glucose-glucose**).

Add the total number of carbons, hydrogens, and oxygens from the two monosaccharides that make up sucrose:

C=                      H=                      O=

What is lost as a by-product during the formation of sucrose from the two monosaccharides (hint: subtract the formula for sucrose from that of the two monosaccharides above)

4. Delete the sucrose molecule and load the starch molecule. This is much smaller than the average starch molecule in plants, so it is easier to visualize.

Which sugar monomer is the building block of starch? (hint: use your drawings from 1 & 2 to help you decide).

How many sugar monomers are present?

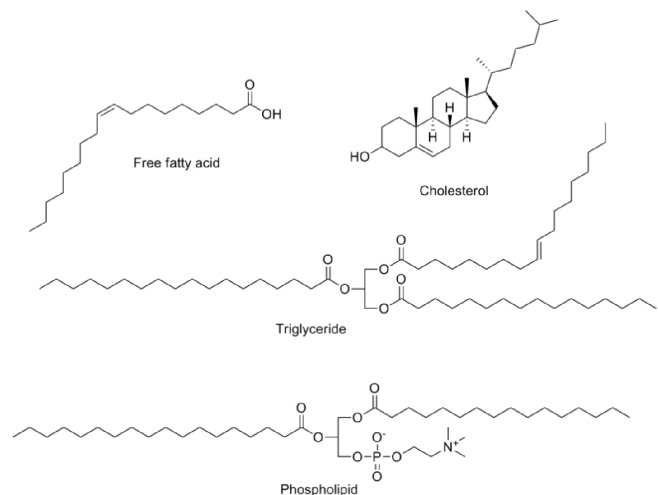
Write the chemical equation for synthesis of this starch molecule below:

## LIPIDS

Lipids are a diverse category of molecules, made up of non-polar molecules, which are unable to be dissolved well or at all in water.

These include many molecules made of fatty acids and steroid-alcohols

such as cholesterol. The diagram on your right shows the structure of cholesterol, a mono-unsaturated fatty acid, a mono-unsaturated triglyceride (fat), and a saturated phospholipid (primary component of the cell membrane).



1. Load the palmitic acid molecule. This is a 16-carbon fatty acid which is a building block of many fats and phospholipids present in plant and animal

cells. Fatty acids all contain a long carbon chain terminated with a carboxylic acid. Draw the palmitic acid molecule below and circle the carboxylic acid group in your structure.

2. Fatty acids can be saturated or unsaturated. An unsaturated fatty acid has a double bond between two of the carbon atoms in its tail. A mono-unsaturated fatty acid has one double bond, while a poly-unsaturated fatty acid has two or more. Which type is palmitic acid, saturated, mono-unsaturated, or poly-unsaturated? (hint: To determine double bond positions, count the number of atoms bound to each carbon. If it is 4, then there cannot be a double bond. The first carbon, bound to the oxygens does not count for this).
3. Delete the palmitic acid molecule and now load the oleic acid. This is an 18-carbon fatty acid that is also common to many fats and phospholipids. Draw the structure of this molecule below. What is the main difference in the structure, other than the length, between oleic acid and palmitic acid? The difference is the result of a carbon-carbon double bond in one of these molecules but not the other.
4. Now delete the oleic acid and load the glycerol molecule. This very simple molecule forms the basis for linking fatty acid chains together into triglycerides (fats) or together with polar and charged “headgroups” in

phospholipids. Draw the glycerol molecule below, making sure to point out the 3 alcohol groups (where the fatty acid chains or headgroups would be attached).

5. Now delete the glycerol molecule and load the POPC molecule. This phospholipid is the most common constituent of cell membranes. You should notice that it is basically just the glycerol molecule you have just looked at, only with two fatty acid chains and one headgroup attached where the alcohol groups were. Draw the molecule, making sure to point out the glycerol unit, the two fatty acids, and the headgroup.
6. The POPC has one saturated and one unsaturated fatty acid chain. In your drawing above, point out which one is unsaturated and indicate the position of the double bond (hint: count the number of atoms bound to each carbon as before – but don't include the carbons bound to oxygen).
7. One of the chains is palmitic acid and the other oleic acid. Use your answers and drawings from 1 & 2 above to determine which is which and label each chain in the above drawing.
8. Which fatty acid has a more kinked structure, the saturated, or unsaturated one?

Extra Credit:

Load the structure of cholesterol into VMD. There are 3 important regions of the cholesterol molecule, the alcohol group, the 4-ring structure that is characteristic



of all steroid hormones, and the hydrocarbon tail. Draw the structure of cholesterol below. Circle and label each of these regions in your drawing.

## NUCLEIC ACIDS

Nucleic acids are the genetic information storing molecules present in every known life form and they are even required for viruses to be infectious. They are made from primary building blocks called nucleotides. The nucleotides with guanine, cytosine, adenine and, thymine bases are found in deoxyribonucleic acid, or DNA. In ribonucleic acid, or RNA, thymine is replaced by another nucleotide called uracil. Each nucleic acid building block also has a sugar and phosphate group attached. This sugar-phosphate-sugar-phosphate sequence is referred to as the “backbone” of the nucleic acid and does not change throughout the molecule. The backbone is different in DNA and RNA, however, since the RNA backbone contains a ribose sugar, while DNA contains a deoxy-ribose sugar. The entire nucleotide-sugar-phosphate unit is referred to as a nucleoside.

1. Load the thymine nucleotide into VMD. This structure contains a phosphate group, a 2-deoxyribose sugar, and the thymine base. Draw this molecule in the space below, labeling these three parts of the DNA building block.
2. Delete the thymine molecule and load the uracil nucleotide into VMD. Draw this structure below, pointing out the phosphate, sugar, and uracil base. Also,

indicate the two major differences between uracil & thymine (hint: one difference is in the base itself and the other is in the sugar).

3. Delete uracil molecule and load the adenosine nucleoside (the nucleotide base + phosphate + sugar). Draw the molecule below and indicate whether it is from DNA or RNA building block. How do you know?
  
4. Delete the adenosine molecule and load the DNA molecule. In the presence of magnesium and calcium, DNA is rarely found as a single polymer strand, but instead forms a so-called double helix structure which is held together by hydrogen bonding between the nucleotides (bases). Each base can only hydrogen bond with its complementary base, so that each strand is an exact complement of the other one. You can imagine how this could prove useful for storing important genetic information, since there is always a backup copy of the code available in the second strand. Use the label function to determine which nucleotides hydrogen bond to which in DNA (just give the two pairs). To enter label mode, press the number 1 on your keyboard. You can then left click on individual atoms in the structure to determine which nucleotide they are from (DT means DNA-thymine, DC means DNA-cytosine, DG means DNA-guanine, and DA means DNA-adenine).

## **C.4 My Favorite Protein**

### **C.4.1 Lesson Plan**

#### *Introduction*

Proteins are a diverse class of biological molecules, which play a crucial role in almost every biological process. Proteins, known as enzymes, are responsible for the catalysis of nearly every reaction in the cell; protein receptors are crucial for the communication of nerve impulses & hormonal signals. The cytoskeleton, which gives the cell its shape and mechanical support, is composed entirely of proteins. One of the most impressive examples of protein diversity are the antibodies and T-cell receptors of our adaptive immune system. These proteins must differ enough that they can respond to all the infectious agents our body will encounter in our lifetime, but they must be identical in their overall 3-dimensional shape in order to perform their job properly. A reasonable question to ask about proteins is: How can they function in so many different ways if they are all made of the same 20 amino acids? Indeed, the basic building blocks of every protein are the 20 essential amino acids listed in your textbook. The difference between individual proteins, however, is the exact number and order of these amino acids. A change at even one location in the sequence of amino acids which make up a protein can lead to a loss of function. This may be because that amino acid was important for the protein to fold up into its proper structure, or it may be because the amino acid was important for reacting with other molecules in the protein's active site.

#### *Lesson Goals:*

3. To improve students' understanding of the relationship between protein structure and function.
4. To improve student understanding of intermolecular forces which stabilize proteins.
5. To help students understand the importance of proteins and the varied roles they play in cells and organisms.

These lesson goals are assessed using the quality of student reports and models. See the grading rubric and handout for details.

*Prior knowledge required:*

We developed this lesson as a follow up to the molecular biology unit in college preparatory biology. We implemented the lesson over two years at Danville and Neuqua Valley High Schools in Illinois. The lessons were taught in introductory, honors and AP classes with a total of 7 different teachers, all of whom were novices with respect to VMD in particular and computer-based molecular visualization in general. Students should understand the following prior to beginning this lesson:

1. Proteins are hetero-polymers of amino acids. There are 20 unique amino acids, some positively charged, some negatively charged, some non-polar, and some polar.
2. A protein's sequence refers to the exact type and order of amino acids. A protein's structure refers to its 3-dimensional form.
3. A protein will fold up into a unique, compact structure, depending on what types of amino acids compose it and the order in which these amino acids are arranged.

4. Hydrophobic amino acids are non-polar, oil-like molecules. They will tend to associate with one-another when dissolved in water. This separation of oil-like and water-like parts of a protein is part of what makes it fold. Hydrophobic amino acids will be found on the inside of a folded protein, unless it is a membrane protein, which will have hydrophobic amino acids on the outside.
5. In addition to hydrophobic effects, hydrogen bonds, ionic bonds (salt bridges), and covalent sulfur-sulfur (disulfide) bonds are important for protein stability.
6. Enzymes are proteins that speed up (catalyze) reactions. They all have little pockets inside them called active sites where the reaction takes place.
7. Protein function is usually very sensitive to structural changes, particularly near an active site or binding site.

*Materials:*

- A computer with VMD installed and an LCD projector for the instructor.
- A computer with VMD installed for each student group.
- The VMD tutorial (“Getting to know RNA Polymerase”).
- A Student handout (“My Favorite Protein”) for each student.
- Data storage is required for saving VMD states or rendered images.

Alternatively, images can be printed, preferably in color.

Software: VMD is freely available at [www.ks.uiuc.edu/vmd](http://www.ks.uiuc.edu/vmd). The software requires 35 megabytes of disk space on a computer running Microsoft Windows 2000, XP, or Vista, or MacOSX 10.3.5 or later.

### *Protocol*

#### *1-5 Days before:*

Prepare students for the activity by going over relevant terms and concepts in the “Prior Knowledge Required” section above.

#### *Day 1*

Introduce the activity and distribute the handout to students. Help them navigate to the PDB website and choose an appropriate protein. Allow them to read a little about their protein before writing down its name and PDB ID on their handout. Lead the class in launching VMD and begin the “Getting to know RNA Polymerase” tutorial. If there is a printer, students can print their “Molecule of The Month” protein handout and read it for homework.

#### *Day 2*

Complete the “VMD Basics” tutorial and help students get their proteins loaded into VMD. They should spend the rest of the day working on creating a VMD representation for their report that highlights the parts of the protein required for its structure and function. Students must save a VMD state or render an image to save their work by the end of the day.

#### *Day 3*

Students continue with their VMD representations and render a final image for their report, which can be saved to a disk, e-mailed or printed if a printer exists. Two to three days are usually sufficient for report writing and building the 3-dimensional model at home.

Note to the instructor:

This lesson can be simplified if the entire class works on the same protein. It may be necessary to warn against plagiarism before report writing begins to ensure protein descriptions aren't lifted directly from the "Molecule of the Month" handouts.

### Grading Rubric

CRITERIA	Poor	Good	Excellent
<p><b>THE VMD MODEL</b></p> <p>(this is worth 25% of the total grade for this project)</p>	<p>Student made an attempt which was significantly below the standard set by his/her classmates and the instructor.</p>	<p>Student made a great attempt at identifying all the applicable features described to the right. Studnet simply missed one or two features that could've been shown.</p>	<p>This model contained where applicable:</p> <ul style="list-style-type: none"> <li>-ID of active/binding site including amino acids involved in the active site chemistry</li> <li>-ID of alpha helices and beta sheets</li> <li>-ID of at least one of the following: disulfide bridges, salt bridges, hydrophobic interactions, or hydrogen bonds.</li> </ul>
<p><b>THE 3-DIMENSIONAL MODEL</b></p> <p>(this is worth 25% of the total grade for this project)</p>	<p>Student made an attempt which was significantly below the standard set by his/her classmates and the instructor.</p>	<p>Student produced a good product, but may have fell short in one of the categories listed below:</p> <ul style="list-style-type: none"> <li>-ID of essential protein features.</li> <li>-Creativity in use of materials.</li> <li>-Simplification of model down to essential features.</li> </ul>	<p>This model contained, where applicable:</p> <ul style="list-style-type: none"> <li>-ID of all the essential features for this protein listed in the box above.</li> <li>-Creative use of materials.</li> <li>-Was simplified to the point that one could really focus on the important aspects of the protein.</li> </ul>
<p><b>THE REPORT</b></p> <p>(this is worth 50% of the total grade for this project)</p>	<p>Student made an attempt which was significantly below the standard set by his/her classmates and the instructor.</p>	<p>Student produced a good report, but fell short in one of the categories listed to the right.</p>	<p>The report included:</p> <ul style="list-style-type: none"> <li>-A concise, clear, and interesting discussion concerning protein function and real-world significance.</li> <li>-A clear and understandable explanation of structural features important for protein stability.</li> <li>- A clear and understandable explanation of structural features important for protein function.</li> <li>-Specific references to important features of the VMD and physical models.</li> <li>-The report was at least 3 pages and was not copied from any website or textbook.</li> </ul>



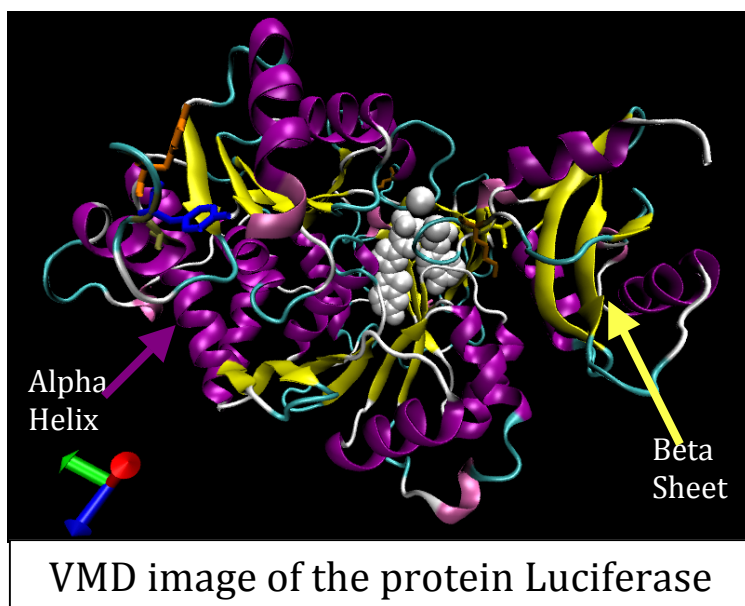
## C.4.2 Student Handout

# MY FAVORITE PROTEIN

### INTRODUCTION:

Proteins are a diverse class of biological molecules, which play a crucial role in almost every biological process. Proteins, known as enzymes, are responsible for the catalysis of nearly every reaction in the cell. You may remember studying the enzymes DNA polymerase, ATP synthase and ribosomes already in this class. Proteins perform widely different tasks in our cells, but they are all made of the same 20 amino acids listed in your textbook. How is this possible? Because a protein's structure - its 3-dimensional form - and its function - how it performs its specific job - are very sensitive to the order of these amino acids in the protein chain. They are so sensitive, that a change (mutation) in the sequence of even one amino acid can lead to a loss of function if that amino acid was important for the protein to fold up into its proper structure or if it was important for reacting with other molecules in the protein's active site.

### WHAT WILL YOU BE DOING:



For this assignment, you will choose a protein and study its 3-dimensional structure using the program, Visual Molecular Dynamics (VMD). You will identify the active site or binding site(s) in your chosen protein and the structural features (e.g. alpha helices, beta sheets, disulfide bridges, & hydrophobic center) which contribute to its stability. You will then highlight these

features using the drawing and coloring methods available in VMD, doing your

best to simplify the rest of the protein so that the important parts can be most easily appreciated.

### **CHOOSE A PROTEIN!!!**

Before we can do anything with VMD, you must select a protein to work with. You may choose any ENZYME (see familiar enzymes from class in the box below). To select your protein, go to **WWW.RCSB.ORG**. This link will take you to the protein data bank website. Scroll down to the bottom of the page to find a special section called "Molecule Of The Month". Click on the "Previous Features" link. This is the list you will select your protein from!

**HOW DO I KNOW WHICH PROTEINS ARE ENZYMES??????????**  
MOST OF THEM ARE. LOOK FOR MOLECULES WHICH END IN ASE,  
IN, OR ZYME. REMEMBER, ENZYMES SPEED UP CHEMICAL  
REACTIONS

### **RESEARCH OF PROTEIN FUNCTION:**

- Click on your selected protein. This will take you to some interesting pages regarding the function of your protein. You will need to report on the following regarding your protein's function:
  - Specifically, what does it do in the cell (or outside the cell)?
  - Why is this protein necessary for survival of a cell and/or a whole organism?
  - ALL of these proteins have something unique and interesting regarding their function, you **MUST** report on the special aspects of this protein's function!

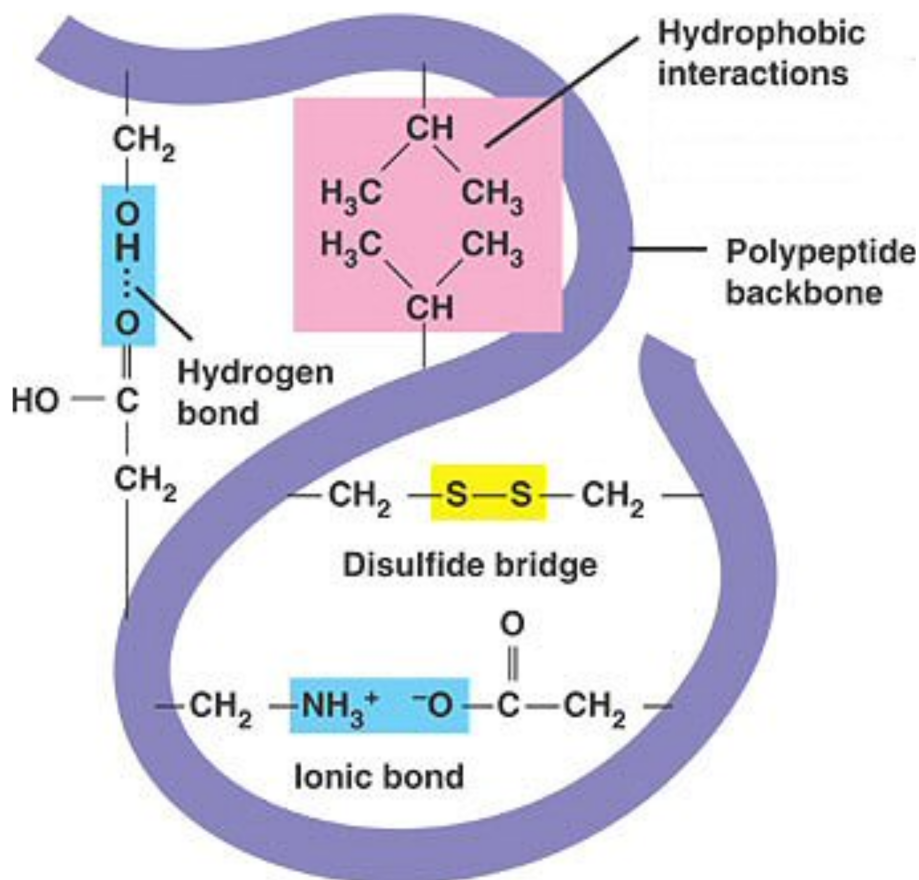
### **FAMILIAR PROTEINS FROM YOUR BIOLOGY CLASS:**

1. **DNA replication** – DNA polymerase, DNA ligase, & nucleosomes
2. **Protein Synthesis** – RNA Polymerase, Transfer RNA, Ribosomal Proteins, & Chaperones of protein folding
3. **Respiration** - Cytochrome C, ATP Synthase, & Cytochrome C Oxidase
4. **Photosynthesis** - Photosystem I & Photosystem II
5. **Cytoskeleton & motor proteins** - Actin, myosin, & kinesin

## RESEARCH OF PROTEIN STRUCTURE:

Your instructor will be leading you through a tutorial which will help you identify important aspects of your protein's structure. You will learn how to use the VMD program in order to highlight the following in your protein:

1. The binding or active site which gives your protein its specific function.
2. Which amino acids are located in your protein's active site.
3. Whether the amino acids in the protein's active site are polar, charged, or hydrophobic.
4. What type of outside molecule binds to your protein's binding/active site.
5. The location of hydrogen bonds, alpha helices, beta sheets, disulfide bridges, salt bridges, or hydrophobic interactions which stabilize your protein's structure (see picture below).



## WHAT YOU WILL TURN IN:

### A PHYSICAL 3-DIMENSIONAL MODEL OF YOUR PROTEIN'S STRUCTURE:

- You will be required to create a 3-dimensional model of your protein using any materials you want from home. Creativity in terms of what materials you use to build your model is encouraged!!
- You may approximate the overall shape of your protein. Do not waste time trying to model every last detail, however, YOU MUST BE ACCURATE, IN MODELLING THE ACTIVE SITE (WHICH AMINO ACIDS ARE INVOLVED IN THE ACTIVE SITE CHEMISTRY?) AND STRUCTURES WHICH STABILIZE YOUR PROTEIN'S STRUCTURE (HYDROGEN BONDS, IONIC BONDS, DISULFIDE BONDS, HYDROPHOBIC INTERACTIONS).

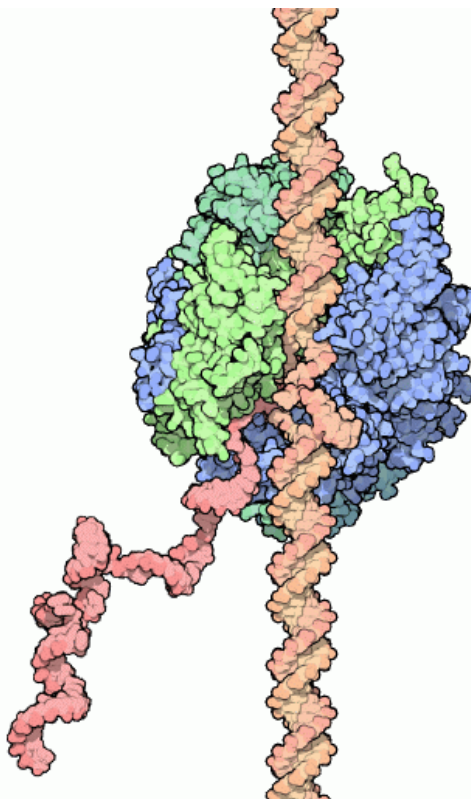
### A WRITTEN REPORT:

- Each of you (not each group) will prepare a written report on your protein. It will be at least 3-pages long, double-spaced. DO NOT COPY FROM ANY WEBSITE OR TEXTBOOK. YOUR REPORT MUST BE ORIGINAL AND MUST CONTAIN ALL OF THE FOLLOWING IN ORDER TO RECEIVE 100%:
  - A concise discussion concerning the function and real-world significance of your protein.
  - At least 1 VMD representation of your protein which highlights interesting features of your protein's structure, such as its binding/active site and features contributing to its stability, like hydrogen bonds (for example in alpha helices & beta sheets), disulfide bridges, salt bridges, or hydrophobic centers. These features must be labeled and discussed in your report!
  - A concise discussion of your 3-dimensional physical model of your protein. You must discuss all of the interesting features you have attempted to model in your creation!
  - Explain at least one feature important for your protein's stability. Exactly how does this make the protein more stable?
  - Explain at least one feature important for your protein's function? Exactly how does this allow your protein to do its job?

### C.4.3 VMD Basics Tutorial

#### Getting to know RNA Polymerase: A Tutorial For Exploring Proteins in VMD

6. You will need to answer some questions and write down some observations during this activity, so please take out a pen or pencil and some paper.
7. Double click on the “VMD” icon in red green and blue (with black background) on your desktop. The window labeled “VMD Main” is the control panel and the window labeled “OpenGL” is where the molecule(s) will be displayed.
8. Select “file”, then “new molecule”. Type in the PDB ID “1msw” and click “load”. Your molecule should be visible now in the OpenGL window (if it is not, see me).
9. In the VMD Main menu, select “Representations” under the Graphics menu. Click the “Create Rep” button in the upper left corner and then type “protein” in the selected atoms field. Now select the “Drawing Method” drop down menu in the bottom left hand corner and select “New Cartoon”.  
10. Now select “Create Rep” again and type “nucleic” into the drop down menu. Choose “VDW” as the drawing method and change the coloring method to “chain”.



11. Highlight the first representation (the selection listed as “all”) in the blue part of the Graphical Representations box and delete it by clicking the “Delete Rep” button in the upper right hand corner.
12. Now let’s go back to the OpenGL screen to have a look at the molecule. You can scale (magnify or shrink) the molecule by pressing the letter “s” on your keyboard and then clicking and dragging the mouse from side to side (try this). You can rotate the molecule by selecting “r” on your keyboard and again clicking with the mouse and moving to one side or the other (try this, too). You can translate the molecule by pressing “t” on your keyboard, and you can choose the center point about which rotation takes place by pressing “c” on your keyboard.

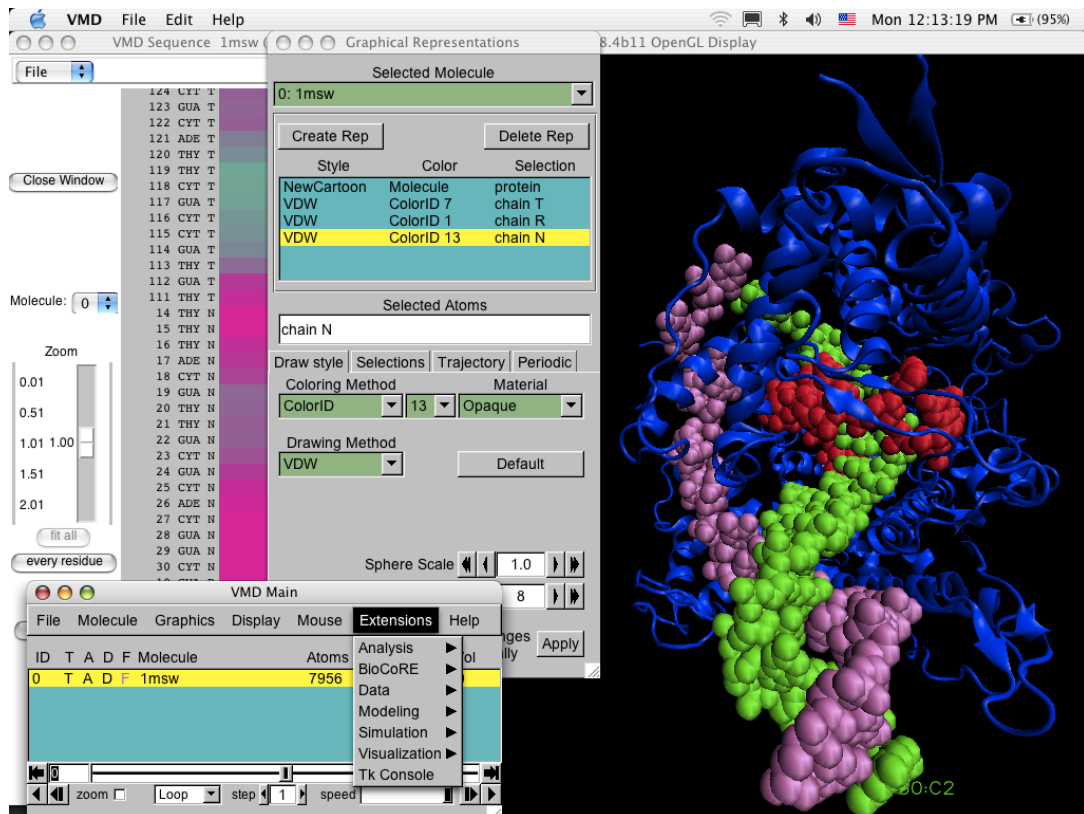
**List of VMD Keyboard Shortcuts (also accessible under the “mouse” menu):**

<b>Key pressed</b>	<b>Action Performed</b>
<b>t, T</b>	Enter translate mode for moving entire molecule
<b>R, R</b>	Enter rotation mode, stop rotation of selected molecule
<b>S, S</b>	Enter scale mode for magnifying or shrinking the selected molecule
<b>c</b>	Choose center about which rotation will take place
<b>x, X</b>	Begin rotation about the x axis, rock back and forth about x axis
<b>y, Y</b>	Begin rotation about the y axis, rock back and forth about y axis
<b>z, Z</b>	Begin rotation about the z axis, rock back and forth about z axis
<b>1</b>	Label atoms selected with mouse left click
<b>2</b>	Enters “bond label” mode, which gives the distance between two atoms selected by successive mouse left clicks

13. The protein is shown in blue. It contains many ribbon-like structures which look like telephone cords – these are called alpha helices- and some others

which look like arrows – these are beta strands. **Take some time to explore the protein structure and write down some of your observations here:**

14. You will also notice in the center of the protein, what look like a bunch of balls in red, pink and green. These are the nucleic acids. More specifically, the pink and green structures each represent two strands of a DNA double helix which has been unwound so that a piece of RNA could be made from this template. The RNA is shown in red. If you haven't already, take some time to read about how RNA Polymerase makes RNA from the PDB website ([www.rcsb.org](http://www.rcsb.org)).



15. Go back to the Graphical Representations window and double click on the representation for protein. This will hide the protein so we can look more

closely at the DNA & RNA. Now click on the “nucleic” representation. Change the coloring method to “ResName”. This will color each of the DNA & RNA bases a different color. Here green corresponds to Cytosine, pink to Alanine, gray to Guanine, and White to Thymine. Can you figure out which of these four bases is absent from the RNA molecule? To help you figure this out, press the number “1” on your keyboard and then click on any atom. You should see a label appear. The label reveals the identity of the atom you clicked. For example, the label A1:O3 means you clicked an Alanine which is the 1<sup>st</sup> amino acid in the protein chain. The “O3” means the specific atom you clicked in the Alanine was Oxygen number 3. Press “r” on your keyboard to exit the label mode.

16. Change the drawing method for this “nucleic” representation back to lines, can you tell which bases are always paired together in the double helix? Spend some time exploring the nucleic acid structures by choosing different coloring and drawing methods and by rotating the molecule around. How far is it across the DNA double helix (how wide is a base pair)? To figure this out, press the number “2” on your keyboard and then click on one side of the helix, then on the other side. You should see a white dotted line appear with a number. That number is the distance in Angstroms – that’s 10 **billion** times smaller than the equivalent distance in meters!! **Write your answer here:**

17. Now you need to save this image. Select “Render” under the File heading in the VMD main menu. Select “Postscript” from the “Render Using” drop down menu. Now select Browse and select the directory where you want to save, then click “Start Rendering”.
18. Now go back to the Graphical Representation menu and the nucleic representation to make it disappear. Now double click on the protein representation. This should cause the protein to reappear in the OpenGL menu – check to make sure this happens. Change the coloring method to



“structure”. What happened? What color are the alpha helices and what color are the beta sheets? **Write your answers here:**

19. Save this image as you did for the nucleic acids. Save it as “yourlastnamePolymeraseCartoon” in your period’s folder.
20. Now go back to the Graphical Representations page and change the coloring method to “ResType”. This will color all positively charged amino acids blue, negatively charged amino acids red, polar amino acids green, and hydrophobic ones white. Now change the drawing method to “Surf”, this will calculate the surface of the protein. It will take some time, so be patient. When you have your protein surface, rotate it around. Does the exterior of the protein appear to be mostly white, or mostly red, green, & blue? **Why do you think that is? Write your answer here:**

## **C.5 Name That Gene, Disease, and Protein**

### **C.5.1 Lesson Plan**

#### *Lesson Goals*

1. To improve students’ understanding of mutations and how a single point mutation in a gene can result in a large phenotypic change like Cystic Fibrosis or Alzheimer’s.
2. To reinforce students’ learning of the Central Dogma of molecular biology, and that genes act by expressing proteins.

3. To reinforce students' understanding that some mutations are more harmful than others, and the most harmful ones will be in a sensitive area of the protein's structure. It will either disrupt the protein's ability to fold (stability) or its function by changing its binding or active site.

### *Prior Knowledge Required*

This lesson is designed as a follow up to the genetics unit covered in weeks 2-4 of the second semester at DHS. Students should understand the central dogma of molecular biology and have a working knowledge of the terms gene, nucleotide, base, base-pair, centromere, chromosome, genome, mutation, genetic disease, genotype, phenotype, bioinformatics, sequence alignment, BLAST and query. In addition, students should understand that a mutation can be a small change in just one nucleotide or the change can involve many nucleotides and that the normal nucleotides can be substituted for something else, deleted altogether, or an extra nucleotide can be added.

### *Definitions:*

- **Gene:** A sequence of DNA that codes for a protein.
- **Nucleotide** (or base): The monomer unit of a DNA polymer, often referred to as a "base". The size of a gene is often given in base pairs (bp) or kilobase pairs (kbp). An average human gene is about 75-100kbp.
- **Gene sequence:** The type & order of nucleotides in a gene.
- **Protein:** A polymer of amino acids which folds up into a compact structure in order complete tasks such as catalyzing reactions.

- **Protein sequence:** The exact type and order of amino acids in a protein polymer. An average protein contains about 300 amino acids.
- **Protein structure:** The 3-dimensional form a protein takes. This shape is entirely dependent on the sequence of amino acids in the protein.
- **Enzyme:** A protein catalyst. It will speed up a chemical reaction without changing the ratio of products to reactants.
- **Active site:** The place on the enzyme where catalysis takes place. The active site is usually in an interior pocket of the enzyme, protected from the environment so that a special kind of chemistry can take place there.
- **Binding site:** Some proteins must bind to another molecule in order to perform their function. The place where the other molecule binds is the binding site. There are usually some special amino acids there which make this site stick to the target molecule.
- **Mutation:** A change in the gene sequence. This can be a small change in just one nucleotide or the change can involve many nucleotides. The normal nucleotides can be substituted for something else, deleted altogether, or an extra nucleotide can be added.
- **Amino acid:** The 20 basic building blocks of proteins.
- **Bioinformatics:** The study of biology using computer and information science. One of the most lively areas of study is the attempt to store & organize the large amount of DNA sequences that have been obtained from organisms from E. coli to humans.
- **Queue:** This is a line. In computer science, it is the line of jobs waiting to be

performed by a computer.

- **Query:** A query is a question. In bioinformatics, though, it is the DNA or protein sequence you input into the computer in order to compare it to the online database of DNA & protein sequences.
- **Hits:** When you submit a query sequence, matches from the database are returned and these are called “hits.”
- **BLAST:** “Basic Local Alignment Search Tool”. This is a search tool that is used to quickly & accurately search online DNA databases for matches against any input sequence.
- **Bit score:** With any BLAST search, hits will have a “bit score”. This is a normalized score that allows you to compare results from multiple searches. The score tells you how well your hit matches your query.
- **Alignment:** This is the process of lining up two or more sequences in order to compare (in a quantitative way) how similar they are.
- **Multiple sequence alignment:** An alignment with more than 2 sequences.
- **Alignment score:** A number that represents how good the alignment is. The score is calculated by adding up every match and subtracting any gap or substitution (non-match) between your query sequence and the hits returned by the BLAST search.
- **Expectation, Expect, or E-Value:** This is the number of different alignments that are expected to occur in a database search by chance, which have the same or better score than your hit. The lower the E value for a hit, the more meaningful the match is likely to be.

- **Chromosome:** The structures in which DNA of higher eukaryotes are found.
- **Autosome:** Any chromosome other than a sex chromosome.
- **Genetic Disease:** A disease that is inherited. It is usually the result of a mutation which causes the protein for which the gene codes to function improperly or not at all.
- **Genome:** The full set of an organisms DNA. For humans, this would include all 21 autosomes and the two sex chromosomes.
- **Centromere:** The structure occurring roughly in the middle of a chromosome that holds individual chromatids together.
- **Protein Data Bank (PDB):** A database of 3-dimensional structures for proteins and nucleic acids.
- **GenBank:** A database of all known nucleotide sequences and their protein translations.

### *Materials*

- A computer with VMD installed and an LCD projector for the instructor.
- A computer with VMD installed for each student group.
- The VMD “Getting to Know RNA Polymerase” tutorial, [VMDBasics.pdf](#).
- Student handout ([NameThatGeneDiseaseProtein.pdf](#)).
- DNA sequences ([DiseaseGeneSequences.pdf](#)).
- Data storage is required for saving VMD images.
- Students will need a pen and paper or word processing program to take notes.

## *Protocol*

(See Wefer et. al<sup>9</sup> or the student handout for a more detailed protocol)

1. Print out the DNA sequences in the quantity required for your class. Each group will need a sequence, but the same sequence can be given to more than one student/group if they are not sitting at adjacent stations. Make sure to cut out the DNA sequence without its corresponding disease name.
2. Lead students to the NCBI website and then to the BLAST page. It is helpful if everyone performs their BLAST searches together, despite the different sequences.
3. Once students find their match, help them navigate to the “Genes and Diseases” online book. Some genes will contain the name of the disease, while others will not. Lead students to the “search book” option until they find their disease.
4. Students should also explore the chromosomes and determine where their gene is located. They will need to take extensive notes during this portion of the activity.
5. Under the main page for each genetic disease is a link to Entrez, which will provide detailed information about the protein and mutations leading to the disease. Students should copy the protein name down, as well as any information about the location of the mutation in the protein sequence and the final folded structure.
6. Help students navigate to the PDB website and search for their disease protein there. They will need to copy down the PDB ID.
7. Help students load their protein into VMD and locate the site of the mutations. In some cases, a mutant form of the protein will be available and that structure can

be compared with the normal version. Otherwise, the student should look to see if the mutation site is near a binding or active site. It could also disrupt stability by interfering with disulfide bonds, hydrophobic interactions, salt bridges or hydrogen bonds.

8. Students will need to render an image and save or print it for their report, which should be turned in a 2-3 days later.

Note to the teacher:

On this assignment it is important to warn students about plagiarism. They typically have difficulty with rephrasing pathology descriptions and may need extra help with this.

*Grading Rubric (continued next page)*

<b>Criteria</b>	<b>Poor</b>	<b>Satisfactory</b>	<b>Excellent</b>
Name and location of gene with explanation of how student found it (10 points)	Name or location is wrong. Methods explanation is weak or not detailed enough.	Name and location are correct. Methods explanation is weak or not detailed enough.	Name and location are correct. Methods explains exactly how student performed the BLAST, how student knew which hit was correct and how student determined for which disease gene was responsible.
Name and description of disease (20 points)	Name is correct but description contains more than 2 sentences which are direct quotes from the "Genes and Disease" online book or another resource are not properly referenced.	Name is correct but description contains one or two sentences which are direct quotes from the "Genes and Disease" online book or another resource are not properly referenced.	Name is correct and description is detailed without copying the "Genes and Disease" online book or any other resource. Any direct quotes are properly referenced.

*Grading Rubric, continued*

<b>Criteria</b>	<b>Poor</b>	<b>Satisfactory</b>	<b>Excellent</b>
Name and normal function of protein (20 points)	Name and function are correct, but insufficiently detailed.  Contains one or more direct quotes that are not properly referenced.	Name and function are correct, but insufficiently detailed.  OR Contains one or more direct quotes that are not properly referenced.	Name and function are correct and sufficiently detailed. Any direct quotes are properly referenced.
Where, in the protein sequence and folded structure, does the mutation occur? (20 points)	Mutation positions are incorrect or missing.	Mutation positions in the protein sequence OR folded structure are incorrect or absent.	Mutation positions are both correct.
Explanation of how the mutation disrupts the protein's function. (15 points)	Explanation is incorrect and not properly explained or infeasible.	Explanation is incorrect, but detailed and plausible. Answer shows the student properly understands how mutations disrupt can protein function.	Explanation is incorrect, but detailed and plausible. Answer shows the student properly understands how mutations can disrupt protein function.
VMD image and description. (15 points)	No image or no description is given.	Image or description does not support the explanation of how the mutation disrupts the protein's function.	Image and description support the explanation of how the mutation disrupts the protein's function.

## **C.5.2 Disease Gene Sequences**

Note to the teacher:



You may add or delete any sequences from this list, but ensure that there exists a high resolution protein structure which can be inspected in VMD. If the mutation causes deletion of an entire protein, this will also not be useful for this assignment where students are asked to inspect protein structures for the site of single point mutations.

### **CANCER**

CAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGG  
GACAGCCAAGTCTGTGACTTGACGTACTCCCCTGCCCTCAACAAGATGTTTT  
GCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCC

### **ALZHEIMER'S**

TGTTTGCGAACTCATCTTCACTGGCACACCGTCGCCAAAGAGACATGCAGT  
GAGAAGAGTACCAACTTGCATGACTACGGCATGTTGCTGCCCTGCGGAATTG  
ACAAGTTCCGAGGGGTAGAGTTTGTGTGTTGCCCACTGGCTGAAGAAAGTGA  
CAATGTGGATTCTGCTGATGCGGAGGAGGATGACTCGGATGTCTGGTGGGGC

### **SICKLE CELL ANEMIA**

GCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAAT  
TCACCCCAACAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAA  
TGCCCTGGCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATT  
AAAGGTTCCTTTGTTCCTAAGTCCAACCTACTAACTGGGGGATATTATGAAG

### **MARFAN SYNDROME**

AGGATCAGTCTGAGACAGAAGCCAATGTGAGTCTTGCAAGTTGGGATGTTGA  
GAAGACAGCCATCTTTGCTTTCAATATTTCCCACGTCAGTAACAAGGTTTCGAA  
TCCTAGAACTCCTTCCAGCTCTTACAACCTCTGACGAATCACAACAGATACTTG  
ATCGAATCTGGAAATGAAGATGGCTTCTTTAAAATCAACCAAAAGGAAGGGA

### **RETINOBLASTOMA**

CTTCTACTCGAACACGAATGCAAAAGCAGAAAATGAATGATAGCATGGATAC  
CTCAAACAAGGAAGAGAAATGAGGATCTCAGGACCTTGGTGGACACTGTGTA  
CACCTCTGGATTCAATTGTCTCTCACAGATGTGACTGTATAACTTTCCCAGGTT  
CTGTTTATGGCCACATTTAATATCTTCAGCTCTTTTTGTGGATATAAAAATGTGC

### **DIABETES**

GCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGCCTGCA  
GCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGC  
TGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAGACGC  
AGCCCGCAGGCAGCCCCACACCCGCCGCCTCCTGCACCGAGAGAGATGGAAT

### **C.5.3 Student Handout**

## **Name That Gene, Disease and Protein!**

### **What You Will Be Doing**

In this assignment you will be introduced to the field of bioinformatics. This is a field of biology that has arisen recently, as more and more organisms' entire genomes are being sequenced and stored into online databases. You will access one of these databases, known as GenBank, to search for the gene that corresponds to a short DNA sequence that we will give you. The genes, you will find, are all associated with genetic diseases, meaning that there is at least one type of mutation in the gene that can lead to a disease in humans.

Once you find the gene that your sequence came from, you will spend time learning more about it and the genetic disease that results from mutations in the gene. You will also find out what protein your gene codes for (what is its name) and where (at which amino acid position) the mutation occurs in this protein. There may be several possible mutations in this gene/protein (as well as other genes/proteins that give rise to this disease). You're required to describe at least one of the mutations in your gene for your report.

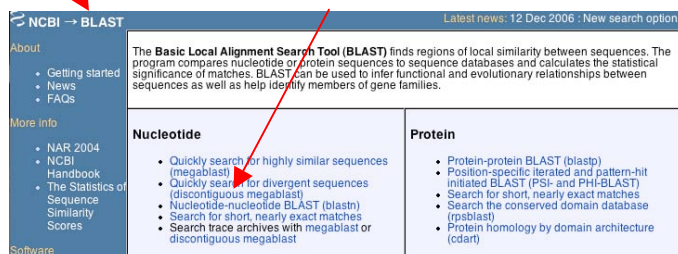
Finally, you will search for the protein your gene codes for in the Protein Data Bank (PDB). There, you will view the structure of your protein and view the site of the mutation. Can you guess why a mutation at that site would disrupt the protein's function?

### Your Report Must Contain The Following:

1. The name of your gene and its location in the human genome (i.e. give the chromosome number and position, such as “very top of the short arm on chromosome 6” or “close to the centromere in the short arm of the x chromosome”).
2. The name and description of the disease associated with mutations in your gene.
3. The name & normal function of the protein your gene codes for and a description of where the mutation occurs (may be more than one place).
4. Describe how the mutation changes your protein’s structure and/or function (i.e. “the mutation is in the active site and replaces a key amino acid involved on the catalysis with one that cannot perform the catalysis”).
5. A VMD image which illustrates number 4.

### PROTOCOL

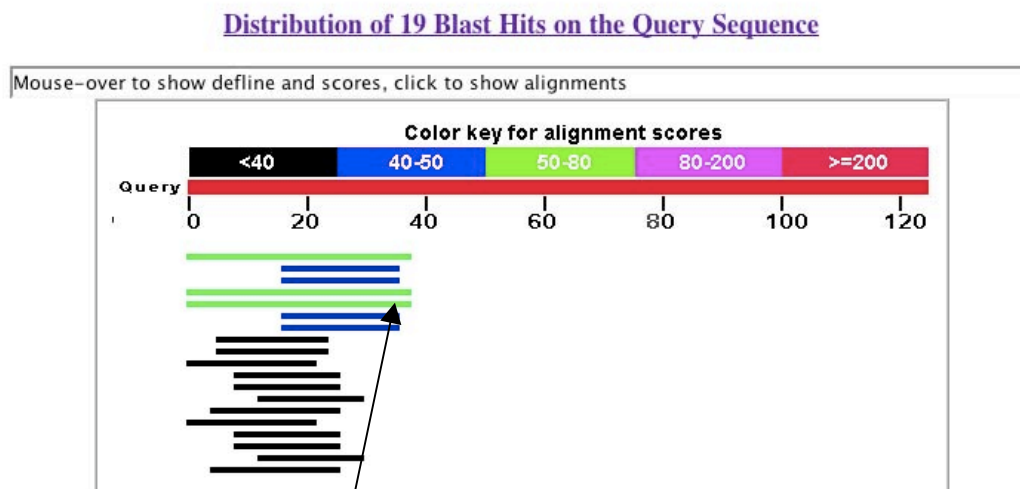
1. Connect to the Internet.
2. Find the home page for NCBI (National Center for Biological Information) at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov).
- 3). Click on the word “**BLAST**,” located on the blue bar at the top of the page.
- 4). Under the heading, “**Nucleotide**”, click on the link, “**Nucleotide-Nucleotide Blast (blastn)**”.
- 5). On the next screen, type (using lower case) the **exact**



nucleotide sequence you were given in the large empty box labeled “search”. Double check your sequence to make sure it is correct, then click “BLAST!”

6). On the next screen, you should receive the message, “Your request has been successfully submitted and put into the Blast Queue”. Wait here a few seconds and then click “**Format!**” bar. If your results are still not ready, you will see a screen asking you to wait for the search to complete. Be patient while formatting takes place.

7). After the search has ended, scroll down to where it says “Distribution of BLAST Hits on the Query Sequence”.



The color key for alignment scores lets you know how well the returned sequences aligned to the one you entered. Your query sequence is listed just beneath this key. It is in red since your query must align perfectly with itself! The other matches will differ from pink (the best) to black (the worst). If you scroll your mouse over these colored lines, you will see the sequence name, the score (“S”), and the expectation value (“E”) appear in the box above the color key. Click on the line with the highest score and the lowest expectation value. This is the sequence that gives the best alignment to your

“query”. From what organism is this sequence? You should be able to determine what genetic disease your sequence corresponds to from the name alone. If not, you will need to do some more digging to find the genetic disease. The alignment of this sequence with your query is shown at the bottom. How many bases are shown in this alignment? How many bases were in the original query (the sequence you originally entered)? Is there a 100% match for the bases which are shown? What is the Score for this alignment (in bits) and what is the Expectation or E-value?

8). Now you will go to the Genes and Diseases part of the NCBI website at

<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=gnd>

Read this main page:

**Genes and Disease** NCBI

Search This book Bookshelf PubMed

Click on a Chromosome View Contents Download PDF Download PDA Browse more Books

**Navigation**

- About this book
- Blood and Lymph Diseases
- Cancers
- The Digestive System
- Ear, Nose, and Throat
- Diseases of the Eye
- Female-Specific Diseases
- Glands and Hormones
- The Heart and Blood Vessels
- Diseases of the Immune System
- Male-Specific Diseases
- Muscle and Bone
- Neonatal Diseases
- The Nervous System
- Nutritional and Metabolic Diseases
- Respiratory Diseases
- Skin and Connective Tissue
- Chromosome Map

**Introduction to Genes and Disease**

*Genes and Disease* is a collection of articles that discuss genes and the diseases that they cause. These genetic disorders are organized by the parts of the body that they affect. As some diseases affect various body systems, they appear in more than one chapter.

With each genetic disorder, the underlying mutation(s) is discussed, along with clinical features and links to key websites. You can browse through the articles online, and you can also download a printable file (PDF) of each chapter.

From *Genes and Disease* you can delve into many online related resources with free and full access. For example, you can visit the human genome to see the location of the genes

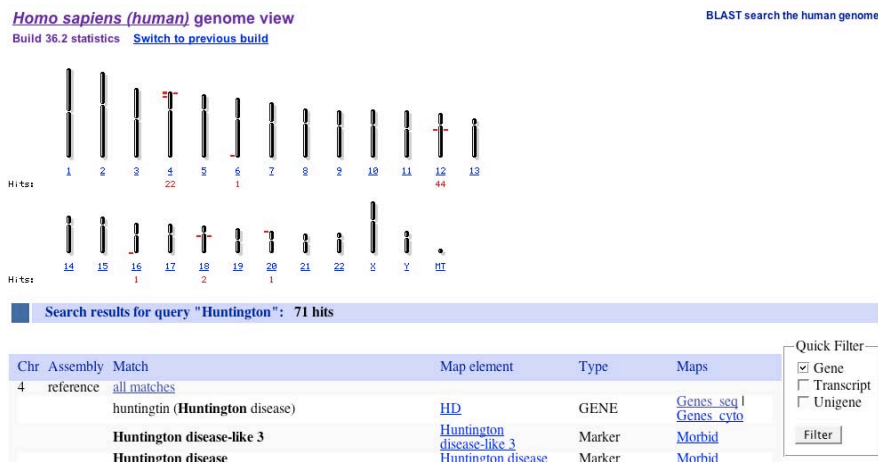
This photograph shows a complete set of chromosomes from an acute promyelocytic leukemia (APL) patient. A new technique called chromosome painting allows visual distinction between chromosomes and can be used to show the chromosome translocations that frequently occur in human cancers. In the case of APL, chromosome 13 is lost, there is a translocation between chromosomes 7 and 15, translocation between chromosomes 11, 15, 17, and between chromosomes 9 and 18. (Look for chromosomes painted with more than one color.) With thanks to Thomas Ried, National Human Genome Research Institute, NIH, for supplying the picture.

9). This site allows you to search genetic diseases by chromosome or by the type

of disease. If you know which chromosome(s) contain mutations associated with your disease, you can click directly on that chromosome (all 23 chromosomes are located at the top of your screen). You can also just click through the various chromosomes until

you find your genetic disease. Also, you can easily find your genetic disease if you know what category it falls under. Is it a type of cancer? Is it a disease of the blood, the eye, the muscle or bone? If you know this information, you can easily click on the corresponding category to get to the page describing your genetic disease.

10). Once you have found the page that describes your disease, you will need to read about it and take notes. What chromosome is the sequence from and what is the name of the gene this sequence is from. What protein does it code for? Find at least one mutation associated with your disease that is not present in the gene you were assigned. To find the additional mutations, select “genome view” in the right panel of this page. You should arrive at a page that shows all 22 autosomes, both x and y sex chromosomes, and the mitochondrial DNA. Red lines on the chromosomes indicate regions of DNA where mutations are associated with your disorder.



You only want the genes relevant to your disease, so find the “Quick Filter” box on the right of your screen. Click “Gene” and then hit

“filter”. You should now see very few (or one) mutation sites in only a few chromosomes. Note which chromosomes (& the region of these chromosomes) are associated with this disease.

11). Now go back to the main description page for your genetic disease. Click the “Entrez Gene” link on the right panel of the page. This will take you to a page containing a list of all genes associated with your disease and their corresponding symbols. Click on these symbols to go to a page which will provide a summary of the function of each gene and the problems associated with mutations. You will need this information for your reports. You will need to find and write down the exact location of your mutation. You may see something such as A83 or Ala83 or Alanine 83. That means the 83<sup>rd</sup> amino acid in your protein is mutated from alanine to something else.

12). Now go to the Protein Data Bank website, [www.rcsb.org](http://www.rcsb.org). Type in your protein’s name in the search bar at the top of the page and press “site search”. If you have any difficulty, try typing just one word (i.e. instead of beta hexosaminidase A, just type hexosaminidase). You will get many “hits”. Click on the one that best matches your protein’s name. Write down the PDB code for this protein, you will need it to load your molecule into VMD.

13). Load VMD from the desktop icon and select “New Molecule” from the File menu in the VMD Main window. Now type in your protein’s PDB code and press enter. The “File Type” should automatically change to “Web PDB Download”. Click “Load” and see that your protein is loaded in the OpenGL window. Spend a few moments refreshing your memory about how to manipulate molecules in VMD. Use the table below as a reference.



<b>Key pressed</b>	<b>Action Performed</b>
<b>t, T</b>	Enter translate mode for moving entire molecule
<b>r, R</b>	Enter rotation mode, stop rotation of selected molecule
<b>S, S</b>	Enter scale mode for magnifying or shrinking the selected molecule
<b>c</b>	Choose center about which rotation will take place
<b>x, X</b>	Begin rotation about the x axis, rock back and forth about x axis
<b>y, Y</b>	Begin rotation about the y axis, rock back and forth about y axis
<b>z, Z</b>	Begin rotation about the z axis, rock back and forth about z axis
<b>1</b>	Label atoms selected with mouse left click
<b>2</b>	Enters “bond label” mode, which gives the distance between two atoms selected by successive mouse left clicks

14). Select “Representations” under the Graphics menu in VMD Main. You will use the selected atoms window to locate your mutations. Select “Create Rep” in the Representations tab and type resid #, where # stands for the amino acid number of your mutation (from step 11). To make it easier to see, change the Drawing Method to “VDW” and the Coloring Method to “ColorID”. If there is more than one chain in your protein, you may want to type “resname NAME and resid #” under selected atoms, where NAME is the 3-letter abbreviation of your amino acid in all caps and # is the resid number. If you still have more than one amino acid highlighted, that’s okay, it means the chains are identical and the mutation occurs in more than one place.

15). Mutations that cause disease occur in parts of the protein structure that are especially sensitive to change, such as active sites and binding sites. Is your mutation in an active site or binding site? To help determine this, first select the first representation and type in “protein” and change the Drawing Method to “New Cartoon”. You can click “Create Rep” again and type in “not protein and not water” to see if your protein has any

co-factors. This is an easy way to locate the active site. Alternatively, if you took good notes in step 11, you may already know how this mutation disrupts your protein and the amino acid numbers of the active site.

16). Once you have an image in VMD that highlights where your mutations occur and why they interfere with its function, you will need to save. From the file menu select “render” and choose render using “snapshot”. Browse to the folder where you want to save and type in “YourMoleculeName”.tga as the filename and click “Start Rendering”.

## **C.6 Proteins as Molecular Clocks**

### **C.6.1 Lesson Plan**

#### *Overview:*

This 1-2 day lesson is designed as a follow up to the evolution unit, which is covered in the last month of the semester at DHS. In this lesson, students explore molecular evidence for evolution at the level of proteins.

#### *Lesson Goals:*

1. Students will learn how multiple protein sequences are aligned and how this alignment is used to build evolutionary trees.
2. Students will understand that different proteins can produce unique trees.

3. Students will understand that regions of high sequence variability (less conserved regions) are usually located away from active sites and the core of a protein structure.

*Prior knowledge required:*

Students should be very familiar with the hierarchy of biological classification and should be comfortable with the terms phylogenetics, cladistics, taxonomy, genetic drift and horizontal gene transfer. It is highly recommended that this activity be preceded by a pen and paper activity using limb morphology or short DNA/protein sequences to build evolutionary trees.

*Definitions:*

- **Cladistics:** the classification of organisms according to their evolutionary relatedness. Relationships are usually represented by a cladogram or evolutionary tree.
- **Phylogenetics:** the quantitative study of the evolutionary relatedness among organisms as determined from morphological and molecular information, i.e. DNA, RNA and protein sequences.
- **Convergent evolution:** this occurs when organisms share traits and/or genes, but are not close relatives. This happens due to genetic drift and horizontal gene transfer.

- **Genetic drift:** a phenomenon that occurs when the same sequence change occurs independently in different lineages because of similar selective pressures or simply by chance.
- **Horizontal gene transfer:** when an organism receives genetic information from a source other than its parents. This occurs only in bacteria via conjugation or transduction by a virus.

*Materials:*

- A computer with VMD installed and Multiseq properly configured. Also, an LCD projector must be connected to the instructor's computer.
- A computer with VMD installed and Multiseq properly configured for each student group.
- Student handout (PhylogeneticsMultiseq.pdf).
- Data storage is required for saving phylogenetic trees (alternatively, they can be copied with pencil and paper).
- Students will need a pen and paper or word processing program to take notes.

*Protocol:*

*1-5 days before*

1. Go over the definitions and terms in the “Prerequisite Knowledge” section.
2. Lead students in a pen and paper activity using limb morphology or short DNA/protein sequences to build evolutionary trees.

### *Day 1*

1. Students should work in groups of 1-3. They will first choose a group of organisms to classify. They will need help choosing a group that is closely related enough, but not too much that no variability will be observed. Popular choices will be placental mammals, carnivores, flowering plants, reptiles, spiders and insects.
2. Students will need to choose 3 proteins common to all their organisms. Again, they will need help with this step as the entire activity will fail if they choose a protein that is not present in all the organisms they want to classify. They will see more or less variability depending on how crucial the protein is for survival (i.e. rhodopsin versus ATP Synthase).
3. Students will need to obtain the established evolutionary relationships from the Tree of Life Web Project at <http://tolweb.org>. They can copy the information on their sheet or print it if a printer is available.
4. The students will then need to obtain the PDB IDs for each of their chosen proteins in at least organism from their taxonomic group. They will head to <http://www.rcsb.org> to search for it by name and copy down PDB IDs for each protein before the end of class.

### *Day 2*

1. Students will begin by loading their proteins into VMD via the automatic “Web PDB Download” feature. They will need a brief refresher on how to load and manipulate proteins in VMD.

2. Next, they will start Multiseq and explore the sequence-structure relationship for their protein. Make sure they understand which chains are protein and which are simply water or other molecules. Give them time to click around and explore their protein with this tool.
3. Now students should choose one sequence or chain from their protein to perform a BLAST search. Their choice will be random for most proteins, though there are some structures in which the importance of chains can be easily discerned because, for example, one chain contains the active site or is in the center of complex where it makes contact with every other chain.
4. After performing the BLAST search, students will need to filter their results. Ensure they have the correct Phylum for their group and that they unselect “all” from the “Domain” and “Kingdom” boxes.
5. Next they will perform their alignments and generate their trees. Encourage students to inspect their alignments, and explain the alignment gaps. Show students how to highlight an entire column, or several columns so those residues are highlighted in the structure in the OpenGL window. They will need to explore their structure to determine which regions have lots of variability and which parts are highly conserved. They will also need to come with an explanation for their observations, i.e. “the conserved region was in the interior of the protein and included amino acids in the active site”.
6. The final step of the activity is to build a sequence-based phylogenetic tree. Students will need to do this for each of their 3 chosen proteins. They can print, save or copy the trees if there is time. They will need all 3 as well as an average

tree for their reports and they will need to explain discrepancies in the trees.

They will need help remembering the reasons for differences in trees generated from single genes/proteins. Acceptable explanations include convergent evolution, differing rates of evolution or because closely related organisms do not have enough differences to produce statistically significant results.

7. The final report should be due in 3 days and should include descriptions of the proteins used for the analysis (what they do and why they were chosen), a brief description of the classified group, including a rooted tree showing the established taxonomic relationships among them, the three phylogenetic trees obtained using Multiseq, as well as an average tree based on these three, explanations of which regions of the protein structure contained the most variability and plausible regions for this variability and explanations for any discrepancies between their three trees.

Note to teacher:

For a more advanced lesson, this activity can be extended to include multiple structural alignments of proteins in addition to sequence alignments. Sequence and structure trees can be compared to determine which alignment method better approximates established taxonomic relationships.

## **C.6.2 Student Handout**

### **Introduction**

In your coursework you have been exploring the classification and phylogeny of species based on morphology (anatomical and fossil evidence). This was the type of evidence that led Charles Darwin to propose his theory of evolution in the 1859 publication “On the Origin of Species.” One of the main predictions set forth by Charles Darwin’s theory of evolution, was that the classifications developed using morphology and geography would also be evident in the genetic material. In other words, the species that were considered to be the most similar anatomically and geographically should also be the most similar when examined at the molecular level. It was not possible to test this prediction during Darwin’s day, but since the invention of rapid methods for DNA sequencing, whole genomes of many species have become readily accessible for comparison. This comparison has confirmed over and over, the predictions of Darwin’s theory of evolution. Organisms that were known to be the most closely related, indeed contained the least number of differences in their DNA. To date, this has provided some of the most convincing evidence for evolution, due to the large volume of available DNA sequences and the extremely low statistical probability of such similarities arising by chance.



## **Assignment Overview**

In this assignment, we will explore molecular evidence for evolution at the level of proteins. You will need to choose 3 proteins to use for this comparison and a group of organisms to classify. You will search multiple genome databases for sequences of this protein in the different organisms you have chosen and then you will perform multiple sequence alignments in order to build phylogenetic trees. You will need to be careful to choose proteins which are common to all of the organisms you want to include in your tree. Also, you will be the most successful if you choose a protein that performs a fundamental task in your organism, so that mutations in this protein are more likely to be lethal. At the end of this exercise, each member of your group will turn in a unique report which includes:

- 1). A description of the proteins used for your analysis (tell what they do and why you chose them).
- 2). A description of each of the organisms classified. Include the established taxonomic relationships for these species in the form of a rooted tree.
- 3). A drawing or picture of all 3 phylogenetic trees you obtained in your analysis as well as an average tree based on these 3.
- 4). A discussion of the similarities and differences among the 4 trees you created and between these trees and the established taxonomy for your group of organisms. What are some possible reasons for the differences you observed? Be as detailed as you can in your explanations.

### **Procedure**

1). First, your group will need to choose a taxonomic group of species to study. For example, you may choose placental mammals, carnivores, flowering plants, reptiles, spiders and insects, single-celled eukaryotes, etc. You may choose any group of organisms which share physical traits and are known to be closely (but not too closely!) related.

2). Next, you will choose at least 3 different proteins which you think will be common to every organism in your group. You can expect to see more or less variability depending on how crucial the protein is for survival (i.e. rhodopsin versus ATP Synthase). To choose an essential protein, think back to those first few weeks of class last semester, what is necessary for something to be considered living? What do all life forms have in common? Now try to think of the proteins that make up those common structures or are responsible for those essential processes.

3). Before you get started, you will need to get the established tree for your set of organisms. To do this, open a web browser and go to: <http://tolweb.org>. Here you will navigate the Tree of Life by starting with the Kingdom and clicking through Phyla, Classes, Orders, etc. until you reach the taxonomic group you are interested in. Copy down that tree below:

4).You will now go to the following website:

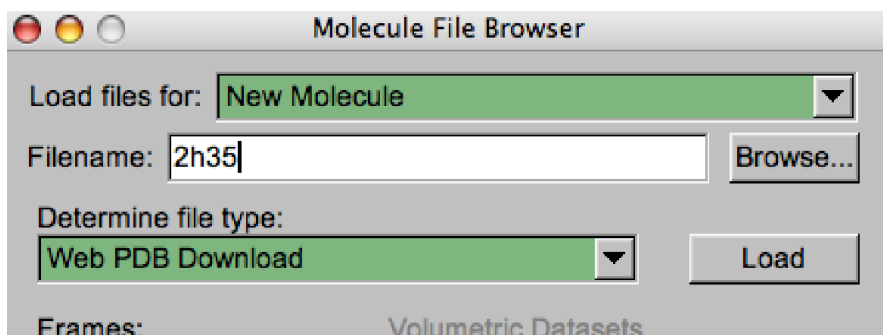
<http://www.ncbi.nlm.nih.gov/>

Type the name of your protein into the search bar at the top. Click on the drop down menu at the left and choose “structure” to find only those proteins which have a solved 3-dimensional structure. Hit “Go” and you should immediately receive a list of proteins. Scroll through the list to find an entry which lists the protein you were searching for. Click on that entry. Note down what is written in the “description” for this entry and note down the code listed to the right of “PDB:”.

The screenshot shows the NCBI Entrez Structure search results for the query 'hemoglobin'. The search was performed using the 'Structure' database. The results are displayed in a table with columns for 'Display', 'Summary', 'Show', 'Send to', and 'Items 1 - 20 of 360'. The results are sorted by 'Summary' and show the first four entries. Each entry includes a checkbox, a PDB ID, a title, a description, and a link to 'VAST, Links'.

Display	Summary	Show	Send to	Items 1 - 20 of 360
<input type="checkbox"/>	1: <a href="#">2H35</a>			Solution Structure Of Human Normal Adult Hemoglobin [mmdbId:42825]
<input type="checkbox"/>	2: <a href="#">2BKM</a>			Crystal Structure Of The Truncated Hemoglobin From Geobacillus Stearothermophilus [mmdbId:42562]
<input type="checkbox"/>	3: <a href="#">2IG3</a>			Crystal Structure Of Group Iii Truncated Hemoglobin From Campylobacter Jejuni [mmdbId:42385]
<input type="checkbox"/>	4: <a href="#">2GLN</a>			Crystal Structure Of Mycobacterium Tuberculosis Trhbn, Gln11ala Mutant [mmdbId:41558]

5). Now you will load VMD. Go to the start menu and select VMD 1.8.5 from the list of programs. Once VMD is loaded, you will need to load your protein. Select “new molecule” from the file menu. Type in the PDB code for your protein (see step 2) and select “Web PDB Download” in the “Determine file type” drop down menu. Now



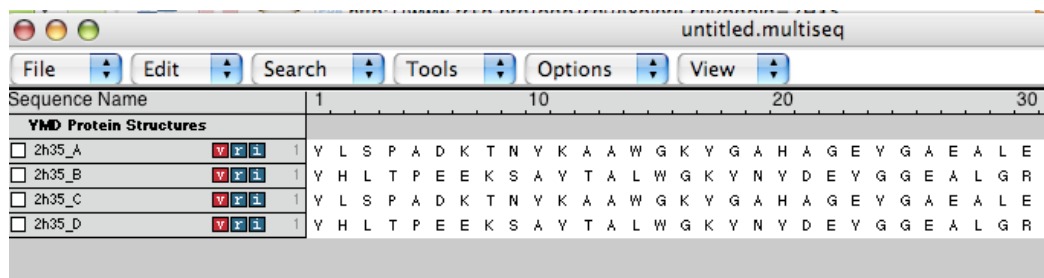
hit “load”. Be patient as your protein structure is downloaded from the web into VMD.

6). Close the molecule file browser window once your protein structure is loaded. Feel free to change the appearance of your protein using the coloring options in the “Representations” window (under the “Graphics” menu). Take a moment to refresh your memory about how to rotate, zoom and label atoms in VMD before proceeding.

## VMD Shortcuts

Key pressed	Action Performed
<b>t, T</b>	Enter translate mode for moving entire molecule
<b>r, R</b>	Enter rotation mode, stop rotation of selected molecule
<b>S, S</b>	Enter scale mode for magnifying or shrinking the selected molecule
<b>c</b>	Choose center about which rotation will take place
<b>x, X</b>	Begin rotation about the x axis, rock back and forth about x axis
<b>y, Y</b>	Begin rotation about the y axis, rock back and forth about y axis
<b>z, Z</b>	Begin rotation about the z axis, rock back and forth about z axis
<b>1</b>	Label atoms selected with mouse left click
<b>2</b>	Enters “bond label” mode, which gives the distance between two atoms selected by successive mouse left clicks

7). Now go to the “Extensions” menu and select “Analysis” then “Multiseq”. This may take some time to load, so be patient. When Multiseq has loaded you will get a window which looks like this:



except your window will contain the sequences from the protein structure you downloaded. Each row lists the primary sequence for one chain of the protein. If there are many chains, you will have lots of rows. If there is only one chain, then you will only have one row. If your structure contains nucleic acids, then their sequences will be listed separately beneath the protein sequences. Also, waters and

ions may be listed in a separate row, with waters labeled “W”. Don’t confuse this with the amino acid tryptophan. You can tell a water row apart from a protein sequence because it will contain the letter W several (probably more than 100) times in a row. If you want to highlight a chain in the structure shown in the OpenGL window, just highlight the row containing its sequence in Multiseq. If you want to highlight a few amino acids, click them one by one while holding down the ctrl key (PC) or command key (MAC). Take a moment to explore your protein’s sequence and structure this way.

8). You will need to select one of the chains from your list to perform a BLAST search – YES, YOU CAN DO THIS RIGHT FROM VMD!!!! It doesn’t matter which chain you select, as long as it is protein and not water or DNA. Select the chain by checking the box to the left of its name.

9). Select “Import Data” from the File menu. Choose the “From Blast Search” and “Marked Sequences” options. Under “database” browse to the “swiss-prot” directory and select the “uniprot\_sprot” database. Now click “OK”.

10). Next you will get a list of results. You will want to filter this list, so you only get results from the organisms you want to classify. Select only the domain, kingdom, and phylum that correspond to your classification group and unselect all from this list and click “Apply Filter”. When you are satisfied with your results, click “Accept”.

11). Next you will want to align your chosen sequences. Make sure that only the sequence you used for your BLAST search and the results that are within the taxonomic group you want to classify are selected. Now select “ClustalW Sequence Alignment” from the Tools menu. In this tab, select “Multiple Alignment” and “Align Marked Sequences”, then click “OK”. You will immediately see that the selected sequences have been shifted around in the Multiseq display.

12). To visualize your results more easily, highlight all the rows of sequences not in your alignment and hit delete. Also, to highlight a particular column, just click the ruler at the top above the column, dragging the mouse along to select the number of columns you want. The dots represent gaps in the alignment. These are not true gaps in the protein, but regions where there may have been an insertion in the gene. After inspecting your alignment, do there appear to be many changes in this protein for your taxonomic group? Are some regions more conserved than others? Highlight the conserved and non-conserved regions in the VMD structure. Where are they located (e.g. on the interior, the outside, nearby the binding site for two chains/subunits, near an active site)?

13). Now you will build a tree with your results. From the Tools menu, choose “Phylogenetic Tree” and select the “Marked Sequences” and “Sequence Tree Using ClustalW” options, then hit “OK”. You will immediately see your tree. Under the “Leaf Text” tab of the “View” menu, unselect everything except “Enzyme Name” and “Species”. Look at the relationships in your tree. Are they what you expected? Are

some organisms classified as more closely related than in the Tree of Life taxonomy?

Why do you think this is? Print an image of your tree by selecting “Save as

Postscript” from the File menu or copy it down here:

14). You will need to repeat steps 4-13 for the two additional proteins from your group. Try your best to select the same organisms in your BLAST search results. It may be impossible to select the exact same set, but do the best you can. Once you have all 3 trees and notes about which parts of each protein had the least and which had the greatest variation in sequence (step 11) you are ready to write your report!!!



## Appendix D

### Computer Codes

#### D.1 Learning Algorithm

```
#include <iostream.h>
#include <stdlib.h>
#include <ctime>
#include <cstdlib>
#include <string.h>
#include <vector>
#include <stdio.h>
#include <limits.h>
#include <time.h>

//User Defined Variable//
#define numParam 3 //Number of parameters for the peptides
int numGood; //Number of Peptides in good data set P
int numBad; //Number of Peptides in bad data set N
int Iterate; //Number of Iterations of the Program
int COUNT;
int Thresh=0; //Error threshold, how large of an error willing to accept

// Functions Start Here
double* add(double y[], double x[]){
double v[numParam];
for(int j=0; j<numParam; j++) v[j]=0;

for(int j=0; j<numParam; j++){
v[j] = (y[j] + x[j]);
}
double *point = NULL;
point = v;
return point;
}

double* subtract(double y[], double x[]){
double v[numParam];
for(int j=0; j<numParam; j++) v[j]=0;

for(int j=0; j<numParam; j++){
v[j] = (y[j] - x[j]);
}
```

```
double *point = NULL;
point = v;
return point;
}
```

```
//Seed Random Number Generator
unsigned time_seed (void)
{
time_t timeval; /* Current time. */
unsigned char *ptr; /* Type punned pointed into timeval. */
unsigned seed; /* Generated seed. */
size_t i;
```

```
timeval = time (NULL);
ptr = (unsigned char *) &timeval;
```

```
seed = 0;
for (i = 0; i < sizeof timeval; i++)
seed = seed * (UCHAR_MAX + 2u) + ptr[i];
```

```
return seed;
}
```

```
/* Main Iterative search for a solution with a given dx, dy, dz
A random vector from data sets P & N is chosen and the dot product
between it and the weight vector is calculated. If the dot product of the
weight vector and the P vector is positive another vecotor is chosen randomly
and the process is repeated. If W dot P is negative, the add function is called.
The dot product of the N vector and weight vector is calculated. If negative
another vector is chosen and the process completed. If positive the subtract
function is called. This is done as many times as the iterations allow.
*/
```

```
double* wvec(double Good[][numParam], double Bad[][numParam], double
Mas[][numParam]){
int i,j,k; //variables used in counting and for loops
double w[numParam]; //the weight vector, defined with the number of parameters being
monitored
double *finalweightvector=NULL;
double dotPrdct_P=0; //dot product of individual peptides in good set P with the weight
vector
double dotPrdct_N=0; //dot product of individual peptides in bad set N with weight
vector
double dotPrdct_M=0; //dot procut of peptides in the master set M with the weight
vector
double *B=NULL;
double *C=NULL;
```

```

int xM=0; //random vector chosen from Master data set
int err = 0; //the error of the weight vector. +1 is given for every misplaced data point
double test3[numParam];
double BestScore=0;
int SizeM=(numGood+numBad);
double TempWeight2[numParam];
double SumDist=0; //Sum of squares of all dot products

for (i=0;i<numParam;i++) {TempWeight2[i]=0;
}
for(i=0;i<numParam;i++) {test3[i]=0;
}
for(i=0;i<numParam;i++) {w[i]=0;
}
finalweightvector = test3;

// **Random generation of the weight vector **
using namespace std;
srand((unsigned)time(0));
//srand((unsigned)time_ seed);
double random_integer;
int lowest=-1, highest=1; // define range of random number
int range=(highest-lowest);
for(i=0;i<(numParam-1);i++){
    random_integer = lowest+double(range*rand()/(RAND_MAX + 1.0));
    w[i] = random_integer;
}

//      cout<<"Initial w = {"<<w[0]<<"< "<<w[1]<<"< "<<w[2]<<"<"}<<endl;
//      cout<<"Score = "<<BestScore<<endl;
//Begin Iterate Function
for(i=0; i<Iterate;i++){
    err=0;
    dotPrdct_M = 0;
    xM=int(SizeM*rand()/(RAND_MAX + 1.0));

//Dot products of the random vectors and the weight vectors are computed
    for(j=0;j<numParam;j++){
        dotPrdct_M += (w[j]*Mas[xM][j]);
    }
//      cout<<"Test Peptide = #"<<xM<<"< "Parameters =
{"<<Mas[xM][0]<<"< "<<Mas[xM][1]<<"< "<<Mas[xM][2]<<"<"}<<"< "Dot Product =
"<<dotPrdct_M<<endl;

//Check if random vector xM is a part of good data set P or bad data set N
/* If the random vector from the Master set is of set 'P' then the dot product

```

should be positive. However if the dot product is less than zero, then that peptide's vector is added to the weight vector, shifting it \*/

```

    if((xM < numGood) && (dotPrdct_M <= 0)){
        B=add(w,Mas[xM]);
        for(k=0; k<numParam; k++){
            w[k] = B[k];
        }
//      cout<<"w after add = {"<<w[0]<<","<<w[1]<<","<<w[2]<<"}"<<endl;
    }

/* If the random vector is from the 'N' set then the dot product should be negative
   However, if it is positive then the data point is allocated to the wrong side and
   is subtracted from the weight vector shifting it to a more proper position */
    if((xM >= numGood) && (dotPrdct_M >= 0)){
        C=subtract(w,Mas[xM]);
        for(k=0; k<numParam; k++){
            w[k] = C[k];
        }
//      cout<<"w after subtract = {"<<w[0]<<","<<w[1]<<","<<w[2]<<"}"<<endl;
    }

/*Dot products for all elements of sets 'P' and 'N' are calculated */
    for(k=0;k<numGood;k++){
        dotPrdct_P = 0;
        for(j=0;j<numParam;j++){
            dotPrdct_P += (w[j]*Good[k][j]);
        }
        if(dotPrdct_P <= 0) err++;
//      cout<<"P dotPrdct = "<<dotPrdct_P<<","error="<<err<<endl;
    }

    for(k=0;k<numBad;k++){
        dotPrdct_N=0;
        for(j=0;j<numParam;j++){
            dotPrdct_N += (w[j]*Bad[k][j]);
        }
        if(dotPrdct_N >= 0) err++;
//      cout<<"N dotPrdct = "<<dotPrdct_N<<","error="<<err<<endl;
    }

//      cout<<"Iteration = "<<i<<endl;
//      cout<<"Current weight = ("<<w[0]<<","<<w[1]<<","<<w[2]<<")"<<endl;
//      cout<<"Error = "<<err<<endl;

```

//If a solution is found, the sum of squares of all dot products is computed and stored in BestScore if it is larger than the existing value.

//The corresponding weight vector, pt, will also be updated with the new value, w, in this case.

```

        SumDist=0;
        for(j=0;j<(numGood+numBad);j++){
            SumDist += Mas[j][0]*Mas[j][0]*w[0]*w[0] + Mas[j][1]*Mas[j][1]*w[1]*w[1] +
            Mas[j][2]*Mas[j][2]*w[2]*w[2];
        }

        if(err <= Thresh && SumDist>BestScore){

//            cout<<"Old Score = "<<BestScore<<endl;
            for(j=0;j<(numGood+numBad);j++) {BestScore = SumDist;
                }
            for(j=0;j<numParam;j++) {TempWeight2[j] = w[j];
                }
            finalweightvector = TempWeight2;

//            cout<<"New Score = "<<BestScore<<endl;
//            cout<<"New weight = {"<<TempWeight2[0]<< ", "<<TempWeight2[1]<< ",
            "<<TempWeight2[2]<<"}"<<endl;

            for(j=0;j<numParam;j++) {w[j]=0;
                }
            // **Random generation of the weight vector **
            using namespace std;
            srand((unsigned)time(0));
            //srand((unsigned)time_ seed);
            double random_integer;
            int lowest=-1, highest=1; // define range of random number
            int range=(highest-lowest);
            for(j=0;j<(numParam-1);j++){
                random_integer = lowest+double(range*rand()/(RAND_MAX + 1.0));
                w[j] = random_integer;
            }

//            cout<<"....."<<endl;
//            cout<<"initial weight vector = {"<<w[0]<< ", "<<w[1]<< ",
            "<<w[2]<<"}"<<endl;

        } //end if
//cout<<"Score 1 = "<<BestScore<<endl;
} //end iterate
//cout<<"Score 2 = "<<BestScore<<endl;
//cout<<"finalweightvector returned from wvec = {"<<finalweightvector[0]<< ",
"<<finalweightvector[1]<< ", "<<finalweightvector[2]<<"}"<<endl;

```

```

//system("PAUSE");
return finalweightvector;
} //end wvec function

int main()
{
system("PAUSE");
cout<<"Number of Good peptides: ";
cin>>numGood;
cout<<"Number of Bad peptides: ";
cin>>numBad;
cout<<"Number of Weight Vector Iterations: ";
cin>>Iterate;
cout<<"Number of Origin Shift Iterations: ";
cin>>COUNT;

// **Declaration of Variables** //
double P[numGood][numParam]; //good data set [num peps], [num params-1]
double N[numBad][numParam]; // bad data set [num peps],[num params-1]
double Master[numGood+numBad][numParam]; //Master data set with Good set P in top
rows and Bad set N in bottom rows
double w[numParam]; //weight vector to be determined. separates the data sets
int i, j, k, t, m, n, row, col, s; //used in for loops for location in vectors
int counter = 0;
double FinalWeight[numParam];
double Hlx5[]={1.086547483,0.42278481,0}; // {Amph-HPMax}
double Booth[]={0.391629281,0.689873418,0};
double TMX3[] = {0.139700174,0.42278481,0};
double Pep60[]={0.114744128,0.458227848,0};
double TMX1[] = {0.291294115,0.649367089,0}; // {Amph-HPMax}
double VPU[]={0.250366462,1.42278481,0};
double Pep11[] = {0.170209935,0.5,0};
double Pep111[] = {0.182488192,0.816455696,0};
double Pep101[] = {0.086547483,0.824050633,0};
double deltax=0;
double deltax=0;
double deltay=0;
double deltaz=0;
double PmaxX=0;
double PmaxY=0;
double PmaxZ=0;
double PminX=0;
double PminY=0;
double PminZ=0;
int MaxX=0;
int MaxY=0;
int MaxZ=0;

```

```

int MinX=0;
int MinY=0;
int MinZ=0;
double FinalScore=0;
double dx,dy,dz;
double dxFinal=0;
double dyFinal=0;
double dzFinal=0;
double PdotProdFinal=0;
double NdotProdFinal=0;
double TempScore;
double *D=NULL; // a pointer is created to bring the weight vector into the main
function
double test2[numParam];
double test1[numParam];
double TempWeight[numParam];
int error1 = 0;
double MasterFinal[numGood+numBad][numParam];

for(s=0;s<(numGood+numBad);s++){
    for(t=0;t<numParam;t++){
        MasterFinal[s][t]=0;
    }
}

for(s=0;s<numParam;s++) test1[s]=0;

for(s=0;s<numParam;s++) FinalWeight[s] = 0;

for(i=0;i<numParam;i++){
    P[0][i] = Hlx5[i]; //Helix5 params are assigned to first row of P
    P[1][i] = Booth[i]; //Booth params assigned to second row
    P[2][i] = TMX3[i]; //TMX3 assigned to 4th row
    P[3][i] = Pep60[i]; //Pep60 params assigned to the third row
}

for(i=0;i<numParam;i++){
    N[0][i] = TMX1[i];
    N[1][i] = VPU[i];
    N[2][i] = Pep11[i];
    N[3][i] = Pep111[i];
    N[4][i] = Pep101[i];
}

for(i=0;i<numGood;i++){
    if(P[i][0]>PmaxX) PmaxX=P[i][0];

```

```

        if(P[i][0]<PminX) PminX=P[i][0];
        if(P[i][1]>PmaxY) PmaxY=P[i][1];
        if(P[i][1]<PminY) PminY=P[i][1];
        if(P[i][2]>PmaxZ) PmaxZ=P[i][2];
        if(P[i][2]<PminZ) PminZ=P[i][2];
    }

    for(i=0;i<numBad;i++){
        if(N[i][0]>PmaxX) PmaxX=N[i][0];
        if(N[i][0]<PminX) PminX=N[i][0];
        if(N[i][1]>PmaxY) PmaxY=N[i][1];
        if(N[i][1]<PminY) PminY=N[i][1];
        if(N[i][2]>PmaxZ) PmaxZ=N[i][2];
        if(N[i][2]<PminZ) PminZ=N[i][2];
    }

    if(PmaxX>0) MaxX=int(0.5+100*PmaxX);
    if(PmaxX<0) MaxX=int(-0.5+100*PmaxX);

    if(PmaxY>0) MaxY=int(0.5+100*PmaxY);
    if(PmaxY<0) MaxY=int(-0.5+100*PmaxY);

    if(PmaxZ>0) MaxZ=int(0.5+100*PmaxZ);
    if(PmaxZ<0) MaxZ=int(-0.5+100*PmaxZ);

    if(PminX>0) MinX=int(0.5+100*PminX);
    if(PminX<0) MinX=int(-0.5+100*PminX);

    if(PminY>0) MinY=int(0.5+100*PminY);
    if(PminY<0) MinY=int(-0.5+100*PminY);

    if(PminZ>0) MinZ=int(0.5+100*PminZ);
    if(PminZ<0) MinZ=int(-0.5+100*PminZ);

    while (counter<COUNT){

        for(deltax=MinX; deltax<MaxX; deltax++){
            dx=0;
            dx+=deltax/100;
            for(deltay=MinY; deltay<MaxY; deltay++){
                dy=0;
                dy+=deltay/100;
                for(deltaz=MinZ; deltaz<(MaxZ+1); deltaz++){
                    dz=0;
                    dz+=deltaz/100;

```



```

    for(i=0;i<numParam;i++){
        P[0][i] = Hlx5[i]; //Helix5 params are assigned to first row of P
        P[1][i] = Booth[i]; //Booth params assigned to second row
        P[2][i] = TMX3[i]; //TMX3 assigned to 4th row
        P[3][i] = Pep60[i]; //Pep60 params assigned to the third row
    }

    for(i=0;i<numParam;i++){
        N[0][i] = TMX1[i];
        N[1][i] = VPU[i];
        N[2][i] = Pep11[i];
        N[3][i] = Pep111[i];
        N[4][i] = Pep101[i];
    }

//cout<<"Peptide Parameters"<<endl;

    for(i=0;i<numGood;i++){
        P[i][0] = (P[i][0]-dx);
        P[i][1] = (P[i][1]-dy);
        P[i][2] = (P[i][2]-dz);
//cout<<"Good Peptide "<<i<<" = {"<<P[i][0]<<","<<P[i][1]<<","
"<<P[i][2]<<"}"<<endl;
    }

    for(i=0;i<numBad;i++){
        N[i][0] = (N[i][0]-dx);
        N[i][1] = (N[i][1]-dy);
        N[i][2] = (N[i][2]-dz);
//cout<<"Bad Peptide "<<i<<" = {"<<N[i][0]<<","<<N[i][1]<<","
"<<N[i][2]<<"}"<<endl;
    }

/**Assign sets P and N to the Master data set for easy access during iteration**/
    for(i=0;i<numGood;i++){
        for(j=0;j<numParam;j++){
            Master[i][j]=P[i][j];
        }
    }
    for(m=numGood;m<(numBad+numGood);m++){
        for(n=0;n<numParam;n++){
            Master[m][n]=N[m-numGood][n];
        }
    }

```

```

//Print out Master set
//for(m=0;m<(numBad+numGood);m++){
//    cout<<"Master set "<<m<<" = {"<<Master[m][0]<<" , "<<Master[m][1]<<" ,
//    "<<Master[m][2]<<"}<<endl;
// }

/*The weight vector is created and a solution is found*/
//D = test1;
D=wvec(P,N,Master);

for(s=0;s<numParam;s++) TempWeight[s]=0;
for(t=0;t<numParam;t++) {TempWeight[t] = D[t];
}
//    cout<<"TempWeight = {"<<TempWeight[0]<<" ,
//    "<<TempWeight[1]<<" , "<<TempWeight[2]<<"}<<endl;

TempScore=0;
for(s=0; s<(numBad+numGood); s++) {
TempScore += (Master[s][0]*Master[s][0]*TempWeight[0]*TempWeight[0] +
Master[s][1]*Master[s][1]*TempWeight[1]*TempWeight[1] +
Master[s][2]*Master[s][2]*TempWeight[2]*TempWeight[2]);
}

if (TempScore>FinalScore){

    FinalScore = TempScore;
    cout<<"Best Score = "<<FinalScore<<endl;

    for(s=0;s<numParam;s++) FinalWeight[s] = TempWeight[s];

    cout<<"Solution = {"<<FinalWeight[0]<<" , "<<FinalWeight[1]<<" ,
"<<FinalWeight[2]<<"}<<endl;

    dxFinal=dx;
    dyFinal=dy;
    dzFinal=dz;

    cout<<"dx = "<<dxFinal<<" , dy = "<<dyFinal<<" , dz = "<<dzFinal<<endl;

    for(s=0;s<(numGood+numBad);s++){
        for(t=0;t<numParam;t++){
            MasterFinal[s][t]=Master[s][t];
        }
    }
}

} //end if

```

```

        }//end dz
    }//end dy
}//end dx

counter++;

}//end while

system("PAUSE");

cout<<"Final WEIGHT VECTOR.....{";
for(i=0; i<numParam; i++) {
    cout<<FinalWeight[i]<<" ";
}
cout<<"}"<<endl;

error1=0;

/*Dot products for all elements of sets 'P' and 'N' are calculated and printed out*/
for(k=0;k<numGood;k++){
    cout<<"Good Peptide "<<k<<" Parameters = {";
    PdotProdFinal=0;
    for(j=0;j<numParam;j++){
        PdotProdFinal += (FinalWeight[j]*MasterFinal[k][j]);
        cout<<MasterFinal[k][j]<<" ";
    }
    cout<<"}"<<endl;
    cout<<"P dotPrdct = "<<PdotProdFinal<<endl;
    if(PdotProdFinal <= 0) error1+=1;
}

for(k=numGood;k<(numGood+numBad);k++){
    cout<<"Bad Peptide "<<k<<" Parameters = {";
    NdotProdFinal=0;
    for(j=0;j<numParam;j++){
        NdotProdFinal += (FinalWeight[j]*MasterFinal[k][j]);
        cout<<MasterFinal[k][j]<<" ";
    }
    cout<<"}"<<endl;
    cout<<"N dotPrdct = "<<NdotProdFinal<<endl;
    if(NdotProdFinal >= 0) error1+=1;
}

cout<<"Sum of squared distances = "<<FinalScore<<endl;
cout<<"dx = "<<dxFinal<<" "<<"dy = "<<dyFinal<<" "<<"dz = "<<dzFinal<<endl;

```

```
cout<<"Error = "<<error1<<endl;  
system("PAUSE");  
return 0;  
}
```